

Europe needs a common market

The collapse of Europe's Athens summit is a calamity but no surprise. The need now is to make the European Community function as originally intended.

EVEN those who expected very little from last week's meeting of the European Communities in Athens are surprised at how little was accomplished. Politicians, but European politicians especially, are almost always at their worst when the defence of national interests has to be conducted in money terms. Last week's summit meeting was one of the twice yearly meetings where, if things are going well, heads of government can congratulate their hosts (by custom, chairman of the Council of Ministers for the previous six months) for a stint well done. Last week, as agreed at Stuttgart in the summer, there was more serious work to do — to bridge the gap between the Community's income and expenditure, already serious and getting worse as the months pass. The failure of officials to reach a workable compromise had been well publicized in advance; the hope last week was that ten successful politicians would be able to pull a rabbit from an apparently empty hat. The casualty of their failure is not so much their own *amour propre* as the opportunity that has been wasted.

The present symptom of trouble is what is called in Europe the Common Agricultural Policy, but that is only the appearance. Not that the appearance is merely a bad joke. The effect of the European policy on agriculture (which covers wine production but not forestry or fish) is that Europe has partly cut itself off agriculturally from the rest of the world by means of import levies, and that the European Commission (which manages the Community) is compelled by law to pay farmers for whatever produce they fail to sell at market prices. Several nonsenses follow. Food prices in Europe are needlessly high, but the Commission cannot control the huge surpluses of perishable commodities it is compelled to own. (A consequential nonsense is that the Commission's programme of research in biotechnology, stillborn at Athens, offered as one of its potential benefits to use the surpluses as raw materials.) A few years ago, the Commission was offering generous pensions to farmers who would give up farming and take their land out of agriculture. Now there is a scheme, called by some the golden hoof-shake, by which a dairy farmer can be paid about £100 for each milking cow by which the size of his herd is reduced.

Even farmers, fearing that quotas on production may be around the corner, are worried by the surpluses — but then produce everything they can, not just for immediate benefit but as a marker for the future. The cost of operating this system, itself more than 70 per cent of the Community's budget, is less than what consumers pay above the odds in needlessly high food prices.

Quotas

Ironically, this curious system had respectable beginnings. The visionaries who wrote the Treaty of Rome in the early 1950s, convinced that the common market they were creating would quickly make the whole of Europe prosperous, argued that farms should not be left out of the general improvement in living standards. The aim was laudable but was entirely uninformed by the great increase of agricultural productivity there has since been, which has made Europe now virtually self-sufficient in most of the commodities it eats. They were also unaware of the difficulties of legislating for constancy of production in the face of the inevitable climatic fluctuations affecting agricultural production. And they reckoned without what has been the chief disappointment of

the past thirty years — the failure of Europe's industrial production to fulfil the optimism of the founding fathers. Yet if Europe were now markedly more prosperous, the cost of its agricultural policy (no different from the cost of farm subsidies in the United States) would seem much smaller.

Quandaries

Faced at Athens with the need for action — to increase the Community's revenue, to reduce the cost of agriculture or both — the ten member governments appear to have forgotten the origins of their dilemma. There will always be food surpluses in Europe so long as Europe is prevented by protectionism from relying on imports from elsewhere to balance deficiencies in its own harvests. So the abolition of the levy on food imports should be a long-term goal — and a reduction of the levies part of any short-term solution. But then there would be a danger that Europe would be saddled not just with its own surpluses but with those from the rest of the world, whence it follows that the prices the Community will pay for surplus produce must also be reduced. But would not that be tantamount to giving up the principle that Europe's farmers deserve to share in the general increase of prosperity? Not necessarily. If it is agreed that farmers deserve special consideration, the best and only logical way to provide it is by means of a sales tax on food consumption.

Such a policy would have the advantage that its cost would be determined by the useful component of farm production, that which sells, and not by the useless surpluses. The yield of a sales tax on food could then be used to ease the transition of farmers now operating well beyond the margins of economic sense into other jobs or, more constructively, into other crops such as trees, on which Europe has no policy worthy of the name. A tax on food would be politically unpopular, but it would cost ordinary Europeans less than the present policy. And it would be a sensible alternative. If the ten governments want to solve the problems left over from Athens, they will have to start thinking about this or something like it. Last week ended in chaos and bad temper chiefly because the warring factions, the net contributors to the

European nomenclature

EUROPE is bedevilled by collaborative organizations. Technically, there are three European Communities — the European Economic Community (EEC, otherwise known as the "Common Market") established by the Treaty of Rome, and two separate (and earlier) communities concerned with coal and steel and with nuclear energy (Euratom). Since 1967, these have been managed by the European Commission, Europe's executive body (civil service) under the direction of a kind of international cabinet (the Council of Ministers) and a parliament (the European Parliament). Since the accession of Greece last year, there are ten members, with two (Portugal and Spain) waiting in the wings.

All members except France and Ireland belong to the North Atlantic Treaty Organization (which also includes Norway and Turkey). The Council of Europe, founded in 1949, has a wider membership (including even Switzerland), the Western European Union a narrower membership (the six founder members of the European Community and the United Kingdom). □



Communities' budget (Britain and West Germany are the only two) and the rest, seem never to have sought agreement on whether farm support in some form should continue.

The governments must now also pay attention to the reasons why the European Common Market has so far failed to yield the prosperity expected 30 years ago. Quite simply, the internal customs tariffs have been abolished but have been replaced by a host of other restraints on trade. (The first consignment of mainland milk imported into Britain after the British Parliament's reluctant acquiescence that imports could not legally be prevented has been held up for the past two weeks at the port of Newhaven because British government inspectors allege that it had been diluted with water.) For months now, some European governments (including the British, to its credit) have been saying that the time has come to make the Common Market work. Only some of them seem fully to understand that success in that direction would help to solve the agricultural question; a true common market in Europe would increase overall prosperity and, if coupled with an effective industrial adjustment programme, would give all ten members a sense of benefit.

The immediate danger is that the Communities will run out of time. There are only six months until the next summit meeting, in France, but the European Parliament may throw a spanner in the works before that by refusing to agree that the Commission should pay the British Government an agreed rebate on its contribution. That would be financially prudent — there is no money — but politically foolish, likely to incite the British Government to an illegal response, withholding part of the monthly payments required of it, for example. The best course, for all who wish to keep the European Communities alive, is to spend the next six months hammering out plans for making Europe more of a reality.

This is the sense in which the Athens meeting was especially calamitous. There has hardly been a time in the past 30 years when so many European governments have had such a direct interest in making Europe work effectively. Belatedly, most governments seem to recognize that they have a great deal to gain from an effective common market. Many of them have been edged in this direction by transatlantic disappointments in the past few months — the persistence of high interest rates in the United States, the failure of the negotiations at Geneva on intermediate-range missiles and the general uneasiness in Europe that its relations with the outside world may be too exclusively determined by what the Soviet Union and the United States are saying (or not saying) to each other. Only last week, the European members of the North Atlantic Treaty Organization appeared to have regained some of the initiative for East-West relations, and most of them seem to be planning to use next month's meeting on European security at Stockholm as a chance to make something of the opportunity. If the Community is not entirely snagged by its money problems, there is a better chance now than for 30 years that the hopes of its founding fathers may be realized.

But what is the chance that in the weeks ahead the whole system will collapse? Many of those who have always doubted the value of the European experiment are still hoping that events will justify their scepticism. They are destined to be proved wrong. For while the members of the Community would no doubt be able even at this late stage to recruit the army of lawyers needed to unravel legal interdependence, events have already taken charge. In spite of the non-tariff restraints too many members of the Community have thrown up, enough commercial partners in different places have been so well cemented together by the trading relationships that have been formed that there would be an enormous outcry if the process were now reversed. Similarly, there are enough long-distance commuters on Europe's railways and airlines to cause a revolution if ever it were seriously suggested that the mobility of labour offered by the Treaty of Rome should be restrained. And what then would happen to joint research projects, too few it is true, that have begun in the past few years? The way things have turned out, the member governments have no choice but to soldier on. They will cause themselves the least trouble if they do so with a good grace. □

Nuclear susceptibilities

The new Labor Government is agonizing needlessly over its policy on uranium.

THE still-new Labor Government in Australia is following a familiar course in trying to evolve a policy on uranium mining: it has sought independent advice from its advisory council, and will thus not have to make up its own mind until next summer. Meanwhile (see page 629), new projects for extracting uranium from the ground will mostly be kept on ice. The chance is only small that the government will be given opinions different from what its predecessors have been told, by Mr Justice Fox's inquiry in the 1960s for example. Australia, by geological accident rich in uranium, has to strike a balance between the commercial exploitation of a valuable asset and its international responsibility to arrange that Australian uranium is not used for making nuclear weapons illicitly. The new government's dilemma is sharper than that afflicting most of its predecessors because of the Labor Party's tendency to talk as if its endowment of uranium were a thorough nuisance — until, that is, it became the government. Now it has a vested interest in a healthy export trade as well.

The choices ahead are simple, and hardly need rediscovery. It is proper that Australia should seek to ensure that its uranium is not used for making weapons illicitly, for which reason it makes sense that uranium should not be sold on the open market as freely as if it were gold or copper. But equally, there is no reason why the government should not allow uranium contracts to be signed at least with the non-nuclear states belonging to the Non-Proliferation Treaty (NPT). For there is now every reason to be confident that the international safeguards system is working well, but it would be counterproductive to follow the Carter Administration's ploy that required buyers of US uranium not to put spent nuclear fuel through a reprocessing plant. The chief difficulty for the Australian Labor Government will arise in its dealings with Britain (which belongs to the NPT but which makes bombs) and France (with China, the only bomb-maker outside the treaty as things are). If the Australian mood is accurately represented by the decision last week not to let a British warship into a dry dock in Sydney Harbour on the suspicion that it was carrying nuclear weapons (depth charges), striking Britain and France off the list of potential purchasers of Australian uranium is probable — but likely to be only a formality, at least while other people's uranium is chasing customers.

The more practical question for the Australian Government is commercial. Although Australia has for more than thirty years supported a programme of research in nuclear energy through its Atomic Energy Commission, and has from time to time flirted with the idea of building a power reactor somewhere, it has won no other direct benefits from this investment. Plainly it is hoping to win some reward from the development of the Synroc process advocated by Professor A.E. Ringwood as the best way of immobilizing fission products in a solid (but it will deserve a hostile response if it tries to make the sale of uranium conditional on the use of that still untested process). What, in its own interest, Australia should be doing is to add extra value to its uranium by building an enrichment plant of some kind, essentially a way of putting onto the world market some of the low-cost electricity still to be had in Australia.

As things are, it is hard to tell whether the Labor Party's squeamishness about uranium will allow the Government of Australia to follow such a sensible course. Unfortunately, the party seems to have fallen into the commonplace trap of confusing the military and civil uses of nuclear fission. Part of the trouble seems to be that Australians' only close experience of nuclear energy has been military — the testing of British nuclear weapons in their own western desert and, more recently, the testing of larger weapons by Britain, France and the United States in the south-west Pacific. The way things have turned out, the Government of Australia could do worse than build a single nuclear power station so as to demonstrate to itself and to its voters that making electricity does not mean making bombs. □

Australian uranium

Government suspends new mining projects

Canberra

THE Australian Government is in two minds on the thorny question of whether or not it should continue to mine and export uranium. It has given the go-ahead for one new mine, at Roxby Downs in South Australia, where gold and copper will be mined together with uranium, which is consistent with Labor Party policy. But that policy also commits the government to phase out the uranium industry, although the Prime Minister, Mr Bob Hawke, together with some members of his cabinet, believes in what is called "sequential development". The result is that eleven other mining proposals, and new contracts on all but one existing mine (Ranger), are in abeyance pending the outcome of a government inquiry to be carried out into Australia's role in the nuclear fuel cycle.

Mr Lionel Bowen, Minister for Trade, says the decision to call a halt to new uranium mining has been taken on economic grounds, because the market cannot sustain new uranium mines. After the federal caucus decision last month, he pointed out in the House of Representatives that 100,000 tonnes of uranium held on the Ranger and Nabarlek mines in the Northern Territory are unsold. There is no processing of uranium apart from its conversion to yellowcake before export, and Australia does not have an indigenous nuclear power industry. In addition, the export of uranium to other countries is subject to bilateral agreements. Australia has negotiated ten uranium treaties, which may limit, for instance, the transfer of nuclear material to a third country.

The federal caucus, essentially the Labor Party's parliamentary policy-making body, has also decided that Australia should review its bilateral agreements so as to strengthen safeguards against reprocessing, against the diversion of nuclear material to countries that have not signed the Non-Proliferation Treaty and against the fabrication of nuclear weapons. It will also examine further opportunities for suppliers of uranium to "advance the cause of nuclear non-proliferation having regard to the policies and practices of recipient countries", and ways in which it can contribute to the safe disposal of nuclear waste.

This may give an impetus to the Synroc project now being developed at the Australian Atomic Energy Commission for the immobilization of high-level wastes from a power reactor. Ironically, in view of the fact that the government has just abolished the Uranium Advisory Council, the caucus recommended the establishment of a commission to consider the "full range of issues relating to the nuclear

fuel cycle".

The study of bilateral agreements and of Australia's role in nuclear non-proliferation and in the safe disposal of high-level wastes has been referred to the Australian Science and Technology Council (ASTEC), an advisory body that reports directly to the Prime Minister. The new chairman of the council is Professor Ralph Slatyer.

That decision is a reaffirmation for the first time since this government took office of the council's importance in decision-

West German biotechnology

Heidelberg grows bigger centre

THE concentration of biomedical institutions at Heidelberg in West Germany continues to grow. The university biology department, now uncomfortably crowded into a single block, will acquire a new institute for molecular biology in 1984. This was started with a grant of DM1 million (£0.25 million) a year from the giant chemical concern BASF. Last week, the company announced that the grant would be extended for a total of ten years.

The federal government will also contribute DM3.5 million a year for five years and the *Land* of Baden-Wurtemberg DM30 million for construction. The BASF money is crucial to the success of the venture, since it gives the institute freedom to compete in the open market for senior scientists by not being bound by the restrictions that impede the mobility of employment in German universities, making it difficult for them to respond to changes in developing fields by hiring non-tenured staff members.

The twin roles of the new institute are basic research and training. BASF will not be entitled to patents or patent licences but instead to staff training. It is also hoped that the institute will help to meet the present shortage of trained biotechnologists in the federal republic.

The new rector at the University of Heidelberg is keen on the development. Professor Giesbert zu Putlitz returned to Heidelberg in October after five years as director of the Heavy Ion Research Institute at Darmstadt, and although himself a professor of physics at Heidelberg, he is well acquainted with biotechnology through his experience as chairman of the National Committee of the Large Research Institutes, which include the institutions at Jülich and Braunschweig.

Dr zu Putlitz would like to see stronger links with industry, but suggests that the institute's main role would be basic research,

making. ASTEC was chosen to carry out the review because it is the body least likely to be accused of bias, but it was nevertheless reluctant to grasp the uranium nettle. Whatever the report says, it is likely to be controversial, and the council's policy of seeking consensus is unlikely to be feasible in this instance.

Many questions hang on the outcome of the inquiry. The present suspension of uranium contracts with France may be reviewed, new contracts for existing mines may be re-examined and conditions for the export of uranium from the Ranger mine to US utilities defined. Most important of all, the Prime Minister, armed with the report, may be able to tell delegates at the Australian Labor Party conference in July next year that the export of uranium is after all not only safe but good for the country.

Vimala Sarma

with final steps towards commercial processes being made in the industrial laboratories. He is chairman of one of the institute's governing bodies and involved in its international scientific advisory board.

A further development at Heidelberg is the formation of a small biotechnology company, Progen Biotechnik, which it is hoped will begin producing antibody-based diagnostic kits and recombinant DNA products in the Heidelberg science park next year. The founders are four senior Heidelberg molecular biologists. Although modest by US standards, the project, which aims at an initial capitalization of DM13 million, is something of an innovation in West Germany and its prospects are by no means clear.

The German business environment discourages venture capital, the source of capital for most start-up companies in the United States. While the business community frequently deplores these restrictions and the federal government is committed to policies favouring their development in West Germany, the legal and financial obstacles may yet prove insurmountable. If Progen Biotechnik gets off the ground, it will be a sign of healthy relaxation.

These new developments give Heidelberg a noticeable edge over other West German centres in biotechnology. The city already provides quarters for the European Molecular Biology Laboratory and for the German Cancer Research Centre, the much troubled national research laboratory with a staff of more than 1,000, now settling down under its new director, Professor G. zu Hausen, who transferred from Freiburg at the beginning of the year.

Apart from the academic departments at the university, Heidelberg also has the Max-Planck institute for medicinal chemistry, a fruitful field for head-hunters in the recent past.

Sarah Tooze

US soldiers' drug use

Dispute about diagnostic testing

Washington

AN Air Force scientist who criticized analytical methods used by the military to detect drug use among servicemen has been removed from his post. Colonel William Manders, chief of the toxicology division at the Armed Forces Institute of Pathology (AFIP), had repeatedly questioned both the validity of test procedures used to detect marijuana metabolites in urine and the quality control in military laboratories.

The testing programme is a major part of the military's effort to crack down on drug use. The Army alone last year analysed more than 100,000 urine specimens, and reported positive results in 17 per cent of them. AFIP is responsible for coordinating quality control procedures and providing technical advice.

The issue is whether the procedure developed by Manders and a colleague, John Whiting, for drug detection is being misused. Its components are radioimmunoassay followed by gas chromatography (GC). In 1981, when the procedure was first introduced, the military's policy was to refer for counselling any serviceman with a positive urinalysis. In December that year, however, the policy changed: a single positive test became grounds for disciplinary action, including court martial.

Manders — supported by at least one prominent civilian toxicologist, Dr Arthur McBay of the University of North Carolina (who is also chief toxicologist for the state medical examiner) — maintains that his test was never intended for use as a forensic tool. A published description of his procedure (*J. analyt. Tox.* 8, 49-52; 1982) notes his laboratory's policy of confirming a positive result from an immunoassay with GC-mass spectrometry — not simply GC as is used by the military now.

Manders' troubles began in March 1983, when he was asked by the prosecution in a court martial at Brooks Air Force Base to testify as to the reliability of the evidence — an immunoassay plus a GC — and he replied that he could not vouch for the reliability of those procedures without mass-spectrometry. At a meeting afterwards with Department of Defense (DoD) officials in charge of the drug testing programme, Manders suggested that if these tests were to be used at courts martial, at least 10 per cent of the GC samples should be verified with mass-spectrometry. But DoD did not change its testing policy.

In October, Manders was called as a defence witness in a court martial at Homestead Air Force Base in Florida, where he repeated his opinion. A few weeks later, he received orders transferring him to Travis Air Force Base in California, to assume the post of officer in charge of the clinical laboratory. Manders had been at AFIP for 12 years, and head of the toxicology division since 1977. The new post

does not involve research and does not require a doctoral degree. Manders has a PhD in biochemistry.

The move is said to have come at the behest of DoD officials upset by Manders' failure to support official DoD policy. An Air Force spokesman said that Manders' transfer became necessary when the Army, which runs AFIP, "released" him from his assignment there, after DoD officials made their wishes known to the Army Surgeon General.

Dr John Johns, the DoD official in charge of the drug testing programme, declined to comment on the case, saying that Manders does not report to him. The situation is complicated by the overlapping responsibilities of DoD and the service branches, and the perennial squabble over territorial rights among them. Each service

France deliberates on bioethics

Paris

ETHICS are not urgent, according to President François Mitterrand of France: "... Give yourselves time: time to reflect, time to discuss, and time to appraise the moral issues", he told the opening meeting of the French national ethics committee for life sciences and health earlier this month. His meaning: better to get it right later than wrong now.

The committee, containing a minority of 15 non-scientists out of a total membership of 36, has already taken nearly a year to assemble, and it will meet every two months for the next year to cope with the backlog of problems placed before it.

A technical subcommittee consisting of eight doctors and four representatives of other disciplines will meet more frequently, starting this month to address questions on medical trials, the medical use of human embryos and the use of surrogate mothers — the creation of "children of two mothers", in the words of committee chairman Professor Jean Bernard. The committee will also consider its own relationship with the media and with similar committees in other countries.

In answer to criticism concerning the dominance of technical scientists in the committee, Bernard has pointed out that since half the committee is to be replaced every two years, nobody is in office for very long.

Each year, the committee will report to the ministers of health and research; it will create a documentation centre on ethics at the medical research council (INSERM) in Paris; and, Bernard announced, since its role is to inform not only scientists but also the general public, it will organize an annual public conference on the issues it tackles, the first to take place next year.

Isabelle Trocheris

has its own testing programme and laboratories, but DoD is responsible for overall coordination of the programme and for quality control, and can decertify laboratories that do not meet uniform standards. AFIP, and Manders in particular, was responsible for providing technical support and advice to DoD.

Johns did say, however, that Manders "has never advised me differently in what we've done". Johns said he had "the utmost confidence" in the use of immunoassays plus GC alone, noting that the cut-off of 100 ng ml⁻¹ in the immunoassay "errs in the favour of the individual". "Our policy is based on the belief that it is valid for forensic purposes." Johns added that his confidence in the GC procedure was reinforced by validation tests with 800 samples using mass-spectrometry. But McBay, who serves on a board of advisers to Johns, said that "the evidence I've seen [from DoD] is that GC alone is not reliable".

Apart from testifying against the prosecution's evidence in a court martial, Manders may also have been a thorn in the side of DoD because of his objections to quality control procedures at the testing laboratories. Manders had in particular criticized the Navy for failing to check a contract laboratory, Mead Compuchem, with external control samples, and then, after this criticism, for sending 400 negative controls in a row after having sent 7,000 test samples without controls.

An Army laboratory at Fort Mead was shut down last month after attorneys preparing a case for nine accused servicemen whose urine samples had been analysed there found evidence said by expert witnesses — including McBay — to show improper interpretation of GCs and contaminated positive and negative standards. Charges against the nine were dropped.

Despite Johns' assertion that GC plus immunoassay is enough, the military seems to be tacitly acknowledging the need for mass spectrometry. According to McBay, "as soon as I get involved in a case, they run a mass-spec".

The problem, McBay says, is that many servicemen faced with a positive urine test (without mass-spectrometry) are disciplined administratively, with loss of promotion, loss of security clearance and the like. "It is ironic, but right now it is almost to the service member's advantage to go to court martial", says Mark Waple, a North Carolina attorney who has represented many defendants in military drug cases. Waple says he has successfully defended 12 servicemen in courts martial by calling on expert witnesses, such as McBay, to repudiate the GC evidence. Waple is also representing 7 servicemen who suffered disciplinary measures on the basis of a single urine sample; he is seeking to expand that to a civil class-action suit on behalf of the thousands dealt a similar punishment since 1981.

Stephen Budiansky

US nuclear power

Disaster by many small cuts

Washington

NOT all the warm-hearted support of the Reagan Administration has stilled the worries of the nuclear power industry in the United States. In recent months, Congress has voted to abandon the Clinch River fast breeder reactor project; a federal grand jury has accused the Metropolitan Edison Company of criminal safety violations just before the 1979 accident at its plant at Three Mile Island; and several new reactors are bogged down in angry disputes about their safety and cost. What follows is a report from some of the trouble spots.

The Shoreham power plant on Long Island seems likely to become the most expensive reactor in the country and may never be turned on. Ordered in 1967, the 820-megawatt plant is expected to cost at least \$3,400 million to complete, 15 times more than planned. The latest in a long series of engineering defects came to light this year with the discovery that all three of the plant's emergency diesel generators were cracked. But attention is now focused on Shoreham's emergency planning procedures, which have turned the troubled reactor into a national test case.

At issue is the insistence of the Nuclear Regulatory Commission (NRC), since the 1979 accident at Three Mile Island, that plans be prepared to evacuate, in an emergency, people living within a ten-mile radius of any nuclear power plant. Shoreham is struggling to comply with this regulation but the local municipality, Suffolk County, is refusing to cooperate. Suffolk argues, with some support from a panel appointed by the state governor, that restricted access roads and unpredictable winds on Long Island make safe evacuation impossible. Unless Shoreham can persuade NRC to approve an emergency plan in which utility employees take the place of municipal workers, the plant may not be allowed to go on to full power.

Shoreham's owner, the Long Island Lighting Company (LILCO), is encouraged by NRC's recent decision to hear details of the company's evacuation plan, but concedes that NRC's decision could go either way. The legal issues are in any case so fuzzy that prolonged litigation is expected to cause further delays in the licensing process. If Suffolk prevails, local municipalities will be able to block the operation of completed reactors simply by refusing to cooperate in emergency planning, a prospect that terrifies the utility. Early hopes that the state government might intervene to end the impasse between Suffolk and LILCO have been dashed by Governor Cuomo's politically astute decision that it would be wrong to force a reactor on a hostile community.

In Ohio, complaints by an employee about the safety of the Zimmer power plant, now upheld by NRC, have

transformed the economics of this 820 megawatt reactor. Although Zimmer is more than 90 per cent completed, the consortium of utilities that owns it believes it may cost as much as has already been spent to rectify structural defects identified by NRC.

The Cincinnati Gas and Electric Company, the major shareholder, has already invested some \$1,600 million in Zimmer, but is considering abandoning the project, even though Ohio law prohibits passing the costs of a cancelled plant on to electricity consumers. Another partner, the Dayton Power and Light Company, is considering a plan to "convert" the unfinished reactor into a gas or coal-fired plant. In California the two reactors at Diablo Canyon have only just received permission to load fuel after a year's delay caused by the last-minute discovery that the reactors, built above the San Andreas fault, would not be able to meet NRC requirements for withstanding earth tremors.

Work on cleaning-up the damaged Unit 2 reactor at Three Mile Island is slowing to a halt because its owner, the General Public Utilities (GPU) Corporation, is short of money. A financial plan drawn up by Pennsylvania's Governor Thornburgh called on the nuclear industry as a whole to contribute \$150 million to the operation,

but pledges from the industry have so far reached only \$66 million.

Adding to the financial squeeze on GPU is NRC's failure to decide whether the company has the management competence and integrity needed to operate a nuclear plant. GPU has responded to allegations that it violated safety procedures by commissioning independent studies that have, so far, exonerated the company. An NRC inquiry, however, found that rules had indeed been violated and one NRC commissioner is refusing to let GPU restart the undamaged Unit 1 reactor until the company's top management is replaced. The company announced last week that its president is stepping down.

Other highly-publicized difficulties have plagued the nuclear industry this year. Twice within a single week in February, a defective circuit breaker prevented the automatic shutdown of a pressurized water reactor in Salem, New Jersey. An equally spectacular financial failure hit the Washington Public Power Supply System later in the year, resulting in the costly abandonment of at least two reactors.

Spokesmen for the industry complain, with justice, that public attention is drawn to the spectacular failures but takes all too little notice of the many reactors that are completed more or less on schedule and which function safely. Even the industry, however, concedes that the fall in the price of oil and slack demand for electricity have plunged the nuclear business into its darkest hour.

Peter David

Comfort for Three Mile Island

Washington

SOME good news, at last, for the General Public Utilities (GPU) Corporation, the luckless owner of the nuclear power plants at Three Mile Island. After an independent investigation, Admiral Hyman Rickover, former head of nuclear propulsion for the US Navy, has declared that GPU has the competence and integrity to run the undamaged Unit 1 reactor. And a separate investigation by a prominent lawyer has dismissed as "unfounded" a number of allegations that GPU has mismanaged the clean-up operation at Unit 2, the reactor seriously damaged in the March 1979 accident.

The reports have come none too soon for the company. In September, an interim report by the Nuclear Regulatory Commission (NRC) claimed that GPU had violated numerous safety regulations in its haste to salvage Unit 2 (see *Nature* 6 October, p.461). At the same time, NRC has withheld permission to restart reactor 1, partly because of doubts about the company's competence. And last month, a federal grand jury indicted the Metropolitan Edison Company, a GPU subsidiary, for "a pattern of criminal violations" in its handling of the plant

before the accident in 1979.

Neither of the new reports will directly affect NRC's decision whether to allow the undamaged Unit 1 to restart, or the continuing NRC investigation of alleged malpractices at Unit 2. But both will go some way towards rehabilitating the utility in the public eye. Admiral Rickover's report, the shorter of the two, is emphatic about the ability of GPU to operate a nuclear power plant. The second report, by Edwin Stier, former head of criminal justice for New Jersey, is more complex.

The 12,000-page report, issued last week, denies claims (upheld by NRC) that the management at Unit 2 has violated safety procedures and (not yet upheld by NRC) that it retaliated against employees who had tried to draw attention to the violations. It concedes that there were "procedural" violations of safety rules but argues that, for the most part, GPU has done its best to strengthen safety procedures. Finally, it rejects — as has NRC — allegations of collusion between GPU and NRC's office at Three Mile Island.

Stier and Rickover conducted their separate investigations at GPU's request, but both authors emphasize their independence.

Peter David

International data flow

Liberals and regulators meet

THE rapid growth in the flow of computer data across international borders was the subject of a meeting of 250 representatives of business corporations and governments from member countries of the Organization for Economic Cooperation and Development (OECD) in London last month.

Transborder Data Flow (TBDF), as it is known, is likely to account for 120 million transactions in 1987, a number comparable with the number of international telephone conversations. Although there is widespread agreement on the need for legislative harmony between countries, the subject has not often been considered at an international forum.

OECD resolutions are not enforceable in law, but do place moral obligations on member governments. No such resolution on TBDF is immediately in prospect, however: last week's symposium was merely to decide an agenda for future discussion.

The focus of attention since OECD's last symposium on the subject seems to have shifted away from privacy protection (where OECD guidelines are held to have been influential in shaping national legislation) to concern over the economic implications of TBDF. International agreement is desirable to eliminate not only technical obstacles but also protectionist tendencies.

Brazil and Canada, for example, have both recently introduced measures to restrict TBDF. The Government of Canada acted to prevent an international bank from transmitting data on accounts to other countries for processing each day, and Brazil prevented an international corporation with a local operation from hooking up to an overseas database.

Mr W. H. Montgomery of the Canadian Department of Communications made his government's case thus: "We would not wish to see data-related activities in the private sector — including planning, financial control, systems design and computer programming — fall to a level in Canada that is low in comparison with our trading partners".

The case for fewer restrictions was made chiefly by Ambassador D. Lady Dougan from the US State Department, who said that the United States "will always insist that the burden of proof is on those who claim that restriction is necessary". According to Ambassador Dougan, "protection of cultural integrity" is an insufficient justification for information control, because it is too often a guise for economic protectionism or censorship of the press. But Mr Y. Utsumi, of the Japanese Ministry of Posts and Telecommunications, warned that the principle of free information flow should not be used to protect the national interests of specific countries.

There is, it seems, no need to encourage companies to make use of the best available technology in the international sphere. But there is as yet no international agreement on how privacy legislation should treat legal entities such as companies. One suggestion made forcefully last week by the Hon. Justice Kirby of the Australian Law Reform Commission was that an OECD committee should be established to tackle legal questions.

There are formidable problems to be faced where criminal acts span several coun-

tries with different legal systems, says Kirby. One recent example concerns a US database, made public under the US Freedom of Information Act, that was interrogated from Norway, where the information obtained was a state secret. Such problems deserve the highest priority, according to Kirby. Existing international copyright agreements are, he says, entirely inadequate to deal with the possibilities that new technology is offering, and there is no consensus on the liabilities of communications operators. The need for agreement on these matters is more urgent now that the means for making international mischief are now to be found in many private citizens' homes. **Tim Beardsley**

Big science

Small not always beautiful

THE question whether Britain can afford "big science" was aired but not settled at a public meeting in London last week organized by the Association of British Science Writers. There was general agreement that more money was required for science as a whole, but participants in the debate were otherwise chiefly seen to be protecting the interests that they represented.

Professor Ken Pounds of the University of Leicester, chairman of the Astronomy Space and Radio Board of the Science and Engineering Research Council (SERC), stressed that in his fields the United Kingdom is at the bottom of the international spending league (0.02 per cent of gross national product, 1 per cent of total UK research and development). With 1,200 people engaged in astronomy and space research in the United Kingdom, the *per capita* spending is £25,000. Yet, he pointed out, such research delivers value for money; although it contributes only 13 per cent of the European Space Agency's budget, the United Kingdom wins 28 per cent of the European Space Agency's support for basic science.

More generally, according to Pounds, geophysics research has been valuable in the understanding of terrestrial climate and resources and astronomy in its capacity to excite and even inspire members of the general public with new concepts such as black holes and gravitational lenses.

According to Professor Derek Colley of the University of Birmingham, chairman of SERC's nuclear physics board, the proportion of SERC's budget spent on the subscription to CERN, the European Organization for Nuclear Research, (£26.5 million in 1982) has decreased since 1979, while answers to questions such as "What is mass?" may have an important impact on society.

The case for "small" science was put last week by Dr Colin Humphries, a metallurgist from Oxford, who produced a vivid comparison of spending in the United States and the United Kingdom on materials

science. The US National Science Foundation, the Department of Energy and the Department of Defense between them spend the equivalent of about £216 million a year on basic materials science, while the United Kingdom spends about £6 million — a 9:1 ratio per head of population. Moreover, as a proportion of SERC expenditure, materials science research has decreased over the past few years. Dr Humphries argued that, while "truth and knowledge" are truly valuable products of "big science", "little science" produces jobs and wealth as well.

This point was also stressed by Professor Mike Hart of King's College, London. While acknowledging the value of big science, he said that small science needs to be done now if "UK Ltd" is to benefit. Would it matter if big science experiments were delayed by a hundred years, he asked? He suggested that the United Kingdom might opt out of experimental particle physics but remain involved on the theoretical side — a notion dismissed as parasitic and impractical by other participants. He made the general point that the science budget, contrary to what the government claims, had not been maintained in the recent past. Part of the problem, he said, is the "sophistication factor" which has significantly increased the costs of measurement devices so that, for example, the installation of the most advanced electron microscope now represents a national undertaking. Even small science, it seems, is big.

In these circumstances, it is not surprising that the participants took refuge in the argument that the total science budget needs to be increased.

Sir David Phillips, chairman of the Advisory Board for the Research Councils, said that the government's response would be scepticism that, within a budget of £500 million from the research councils and a similar amount from the University Grants Committee, a satisfactory balance could not be achieved. **Philip Campbell**

Italy in space

Minister Granelli looks ahead

THE Italian national space programme — a gigantic effort by Italian standards, designed to give Italy a place in the world satellite systems market — is endangered by swingeing cuts in the government budget, science minister Luigi Granelli said in London last week. Granelli was in passage between the launch of Spacelab at Cape Canaveral in Florida and the Italian space base of Malindi in equatorial Kenya.

Although a lawyer, Granelli has clearly been bitten hard by the space bug. "I am determined to intervene heavily to allow investment in this sector to go on", he said. "Our space plan has been so positive, we feel that a national space agency ought to be established."

But the minister of the treasury clearly feels otherwise: of a 119,000 million lira (£50 million) space budget this year (1983), 69,000 million lira have been withheld. Granelli hopes to win this back in the 1984 budget, but it may be two or three months before he knows whether he has succeeded.

Meanwhile, then, the programme must be delayed, putting a brake on ITALSAT (the biggest project), a communications satellite due for launch in 1987; IRIS, an upper stage for the US space shuttle, capable of taking a 900-kg satellite into geostationary orbit; TSS, a "tethered satellite", for aeronomy, which would physically hang on a wire 100 km below the shuttle, dragging through the upper levels of the atmosphere; and SAX, a scientific X-ray satellite. "There are programmes under way that cannot be interrupted", said Granelli. "Cuts make the existing programmes useless."

Italian industry is "very interested" in space, and even in the use of Spacelab, Granelli claimed. Italy had put up 18 per cent of the original cost of Spacelab (about 14 per cent of its final, \$800 million price tag), and this was the largest sum outside West Germany, the prime constructors. The work had involved Italian industry in some "very advanced technologies" and generated important collaborations with US companies. "It's impressive", Granelli says. And he claims that although the Spacelab construction was supported by the government, Italian companies are now prepared to take advantage, on their own behalf, of opportunities for experimenting in microgravity.

Granelli admitted, however, that industrial interest in this "unique" research had not been as strong as hoped, but argued that it will increase. "The same thing happened with Ariane and the shuttle. Now there's a waiting list. Costs will fall. And we can send up unmanned experiments."

Italy is also interested in working on a space station, and on expanding its Kenyan base, according to Granelli. Kenya could be an important launch pad because its

equatorial position yields a greater payload for the same launcher, when compared with Cape Canaveral and even the European Space Agency's base at Kourou in French Guiana. Negotiations are under way "to improve links" with Kenya.

On other Italian matters, Granelli gave little hope of a reform of the National

Acid rain

What cost, what benefit?

As pressure everywhere increases for action to reduce atmospheric emissions of sulphur dioxide and nitrogen oxides, the British Government has reaffirmed its view that more research is still the most responsible course of action. In a briefing document, the Department of the Environment acknowledges that it is "generally accepted" that acid deposition has contributed to the acidification of lakes and streams in sensitive areas, but says there is "still little evidence" that long-range transport of acid pollutants has caused damage to forests such as that reported from West Germany. Meanwhile, the British Government has spent £1 million on research on the subject this year and hopes to spend more in 1984.

Some of the problems were aired at a meeting earlier this month in London of the Watt Committee on Energy. Sir John Mason, who is directing a study for the Royal Society in collaboration with the Royal Swedish Academy and the Norwegian Academy of Sciences and Letters, said that "representivity of measurement" would be a major problem. His

UK private research

At the height of the recession in 1981, industrial research and development in British industry seemed to be modestly more vigorous. But the total numbers employed in private industry, 175,000, had decreased by 18,000 since 1969, with savings on administrative and clerical staff and an increase in the numbers of scientists and engineers in research and development laboratories from 59,000 to 67,000.

These are among the results of a survey carried out by the Department of Trade and Industry and published last week in the department's house journal British Business.

The cost of intramural research has been remarkably constant for nearly two decades. In real terms (constant-value money) intramural industrial research cost £1,400 million in 1964 and £1,661 million in 1981, according to the results of the survey. But spending in 1981 (34 per cent of which appears to have been financed by the British Government) was higher than in any previous year. □

Research Council (CNR), whose inflexible, bureaucratic proportions system — based on the civil service statutes — strangles proper research management in Italy; but he announced that he was preparing a white paper on scientific research, to be published in 1984, identifying areas of waste. The Italian research and development budget might be little more than 1 per cent of gross national product but it could be more efficiently spent, Granelli argues.

Robert Walgate

study will concentrate on the chemistry of rainwater as it percolates through soil.

Just how acid rain kills trees is still in doubt. The Ulrich hypothesis that the toxicity of aluminium mobilized from the soil is followed by pathogenic action on roots seems to be falling out of favour. Attention now seems to centre on ozone peaks as a possible cause of forest damage, probably in association with other factors. Uncertainties about the mechanisms of chemical conversion and damage are such that the benefits of a substantial reduction of sulphur dioxide emissions in Europe cannot be predicted. Cost-benefit studies are therefore impossible.

Sir Walter Marshall, chairman of the Central Electricity Generating Board, agreed. If it could be shown that to halve the board's sulphur dioxide emissions would actually cure the problem in the most cost-effective way, he said, then "we should get on with it". The board is responsible for more than half the British output of sulphur dioxide and, together with the National Coal Board, has provided £5 million to support the Royal Society study. As things are, Sir Walter said, there seems a good chance that a large investment to reduce emissions might produce negligible benefit. While nothing is settled, it seems agreed that a 30 per cent reduction of British emission would cost about £1,000 million.

The governments of West Germany, Norway and Sweden take a more positive line: West Germany has introduced strong measures to reduce sulphur dioxide and nitrogen oxide emissions and in Norway and Sweden pamphlets are distributed to tourists saying "our lakes are as clean as your industry".

The European Economic Community is, by its own standards, moving rapidly towards the control of acid emissions, although agreement on an enforceable directive fixing emission limits for industrial plants is a long way off, given industrial arguments that enforcement by emission standards may be needlessly expensive. More recent proposals that all member states should reduce sulphur dioxide emissions by the same proportion (one target is 50–60 per cent by 1995) seem likely to be more acceptable.

Tim Beardsley

UK nuclear power

A time for boosting morale

THE British Nuclear Energy Society and the British Nuclear Forum last week heard Mr Giles Shaw, Under Secretary of State for Energy, loudly sing the praises of nuclear power as a "vital contribution to our national energy supply".

In normal circumstances it might be thought that that hardly needed saying, but Mr Shaw's speech was not made in normal circumstances. It was an attempt to restore confidence to an industry that has been shocked and hurt by the recent critical press attention over the radioactive sea discharges from the fuel reprocessing plant at Sellafield (formerly Windscale) in Cumbria, operated by British Nuclear Fuels Limited (BNFL).

A self-perpetuating circle of mistrust seems to have grown up between the press and BNFL (and, by association, the whole of the nuclear power industry). A television programme broadcast in early November that raised the possibility of a link between the discharges and a locally high incidence of childhood leukaemia raised extensive interest and prompted a government inquiry. Then, only two weeks later, a slick of chemical solvent containing 600 curies of beta radioactivity was washed up onto the beach near the discharge pipeline.

The mishap appears to have been relatively unimportant from the viewpoint of radiological protection but it was a public relations disaster. A section of public beach had to be closed temporarily and the Department of the Environment issued a warning to the public to avoid "unnecessary use" of the beach. The department has been removing pieces of debris that have been washed up with, it says, activities up to 1,000 times background. BNFL said that contaminated seaweed found on the beach could be kept in contact with the skin for "very many hours without ill effect", which probably did not have the desired impact.

The liquid discharges from BNFL's reprocessing plant are the largest routine discharges of radioactivity into the environment in Britain, and possibly in Europe. The government departments that authorize the discharges have always maintained that radiation doses to the most affected individuals are well within acceptable limits laid down by the International Commission on Radiological Protection and the National Radiological Protection Board (NRPB), but the safety margins shrank rapidly last year when a revised assessment of shellfish consumption and new experimental evidence on the uptake of plutonium from the gut together increased estimated doses from plutonium by a factor of fifteen. As a result, dose to the most exposed group in 1981 was estimated to be 69 per cent of the internationally agreed limit for members of the public and well over the recommended annual limits

for prolonged exposure.

The Ministry of Agriculture, Fisheries and Food says it has not attempted to apply the new gut factors to consumption in the 1970s, when discharges of alpha activity were five times higher: it says that to do so would be "unhelpful". The main objective is to reduce discharges in future (although not in the manner attempted by Greenpeace, a conservation group, before a court injunction ordered it to desist from interfering with the discharge pipeline). About £100 million is being spent on reducing emissions, and discharge procedures have been tightened to prevent a recurrence of last month's events.

The United Kingdom Atomic Energy Authority is acutely conscious of the damaging effect of such publicity, and intends to launch a public relations offensive in the new year that will involve "admitting where we have made mistakes". Mr Clifford Blumfield, director of the Dounreay fast reactor development centre in Caithness, Scotland, has recently written to all employees at the centre and their families saying that discharges from the Dounreay

site are a small fraction of authorized limits and that the effects of Dounreay are only just detectable above background levels, and then only close to the site. To the management of BNFL this must look like cutting and running.

Sir Hermann Bondi, chairman of the Natural Environment Research Council, told last week's gathering that the "cheeseparing attitudes to capital investment in the 1960s" that had resulted in "technically totally unnecessary discharges" may yet cost the industry dear. The development of nuclear power was, he said, essential, not for Britain — "I couldn't care less if we have to pay a few pence more for electricity" — but for the developing world, which would suffer badly if Western reliance on oil as a fuel forced up prices still further.

This argument in fact contrasted rather sharply with one made later by Mr Shaw, who said that nuclear power might reduce generation costs and hence prices to the consumer: "The French have recognized the importance of nuclear power to their economy . . . and not surprisingly they also have cheaper electricity". But in the general air of bonhomie that was the real reason for the gathering, nobody seemed to notice.

Tim Beardsley

UK student numbers

Government predictions awry

THE Department of Education and Science has got its sums wrong on future student numbers in the United Kingdom, according to the Association of University Teachers (AUT). The department's figures fail to take into account changes in social class composition and the increasing numbers of women entrants into universities, says AUT, and could mean that "thousands of youngsters who will be well qualified to enter higher education will not be able to obtain a place".

One critical unknown in determining future student demand is the age participation rate, or the proportion of the population at any age taking up a place in higher education. The size of the 18-year-old population is, of course, known up to the year 2000. Consequently all projections of future student demand hinge on assumptions about the trend in future age participation rate of different groups.

The 18-year-old population peaks this academic year and next. The government predictions of demand for all full-time education assume an age participation rate that increases steadily and results in a peak for demand this year and next year (at around 165,000 entrants) and then drops to about 120,000 by the mid-1990s. The Royal Society has produced estimates that broadly agree with these.

The Committee of Vice-Chancellors and Principals has also produced projections of the number of university students, and

these show no fall from the present peak during this decade and only a modest fall thereafter. The reason for the difference is that the committee assumes that the age participation rate for university entrance will increase more rapidly than for education as a whole.

The latest figures from AUT take this assumption even further, extrapolating existing trends in the age participation rate of women and in the social class composition (most entrants to higher education come from social classes 1 and 2). On the basis that parity in male and female participation will have been reached by the early 1990s and that the number in social classes 1 and 2 will increase by about half by 1998, AUT says that demand will increase until the 1990s before falling back to present levels. It argues that the skilled manpower needs of the economy in the year 2000 will be such that the government should plan for higher education on the assumption that the trends will continue. "It is philistine folly to cut places in future in order to save money now", says AUT's general secretary Diana Warwick.

What does the Department of Education and Science have to say about AUT's calculations? "Well, you know how it is with statistics", said a spokeswoman. "You can make them show anything you want. But our statisticians will be looking at the AUT figures to determine the areas of difference."

Tim Beardsley

Arms embargoes

Contraband cargo upsets Sweden

Stockholm

THE discovery in Helsingborg last month of a shipment of what appeared to be strategically sensitive material allegedly from the United States via South Africa but destined for the Soviet Union has caused considerable embarrassment in Sweden, which formally maintains a policy of strict neutrality to East and West. Coming, moreover, at a time when the Swedish Government was preparing a declaration of its intention to shoot down any cruise-type missiles infringing on its airspace, the discovery of the Helsingborg hardware, and the subsequent finds of software and further hardware at Stockholm's Arlanda airport and in Malmo, became a matter of unusual delicacy.

The Ministry of Foreign Trade is extremely sensitive about the issue. Members of the Swedish Parliament, when questioned, say they know no more than has appeared in the press. The Stockholm International Peace Research Institute (SIPRI), which is supported by the Swedish Government, was not consulted, although it supports full-time research on problems of the transfer of strategic material. Little more hard information has officially emerged than that some 25 tonnes of suspect material have so far been discovered. An unofficial source in Helsingborg has stated, however, that the finds there included 28 minicomputers of type PDP 11/03 and one of the type PDP 11/23 built by the Digital Equipment Corporation (in the United States), two personal computers of type DSD 8080 made in South Africa, and that the Malmo cache included French computer equipment made under licence in Hong Kong. Among material confiscated in Hamburg, which had formed part of the original consignment, was a VAX/11/782 computer also from the Digital Equipment Corporation.

In the circumstances, the Swedish media have concentrated on the more exotic details of the story. The owner of the Helsingborg containers was identified as Richard Muller, a citizen of the Federal German Republic, who has allegedly been blacklisted by the United States since August 1981 on the grounds of suspected technology transfer to the Soviet Union, and who is said to own a summer villa remarkably close to the Musko secret naval base. More to the point, Herr Muller undoubtedly owns two electronics firms: Semitronik AG in Jesterberg-Wiedenhof and Integrated Time AG in Malers, near Lucerne. According to a Swedish lawyer, Claes Broman, who has been retained by Integrated Time since mid-November, the suspect equipment was intended to be used in joint ventures with one or more (as yet unidentified) Swedish companies.

The South African end, moreover, is alleged to have links with Dieter Gerhart,

who was recently sentenced by a South African court to 20 years imprisonment on charges of espionage for the Soviet Union.

So far, the Swedish Foreign Trade Minister, Mats Hellstrom, has said only that Sweden "actively participates" in the United Nations embargo on the sale of weapons to South Africa, and that the import of such material into Sweden would require a special cabinet order. (In fact, such a law was passed on 22 November, apparently in response to the first Helsingborg find, and made retroactive to cover material which had already arrived in Sweden.) The suspect containers, Mr Hellstrom continued, made the whole question of control over strategic material passing through Sweden one of pressing importance. But he has not committed himself on whether the disputed containers did, in fact, contain strategically sensitive material — this decision has been left to a team of "experts" whose report is imminent.

As a neutral country, Sweden has not acceded to the CoCom embargo on the West-to-East transfer of strategic material. The nature of the impounded hardware and software is, therefore, hard to determine — if the Swedes were simply to check it against the CoCom list, they could be accused of a pro-Western bias — particularly after a meeting on 28 November of the head of export control of the United States Trade Department, Theodore Wu, with representatives of the Swedish export industry, in which he urged them to support the United States viewpoint on the dangers of technology transfer to the Soviet Union. Equally, if the Ministry of Foreign Trade decides that it is, indeed, sensitive material, simply to return it to the United States would not necessarily block the use of Sweden as a transit point for the shipment of other sensitive computer material to the Soviet Union.

A further complication is the widespread rumour that the discoveries have been somehow engineered by an unnamed foreign power to imperil the next round of "Helsinki" meetings on security and cooperation in Europe planned for Stockholm next month. Whatever the decision of Mr Hellstrom's experts, it is likely to evoke adverse public comment.

Even so, the incident has revealed a major deficiency in the international understanding of arms control. Definitions of what constitutes strategic material have so far been left to the military alliances and potential belligerents. For neutral countries such as Sweden, which may be unwillingly compromised by attempts to circumvent the embargoes imposed by the major powers, there are, as yet, no accepted guidelines.

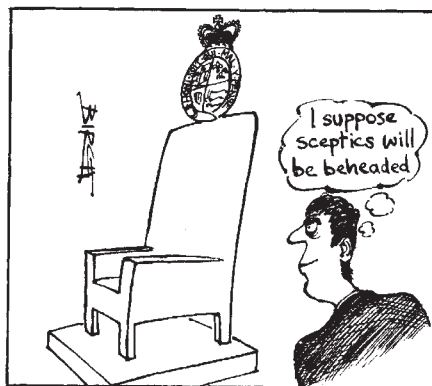
Vera Rich

Paranormality

UK universities vie for funds

RESEARCH into parapsychology may be institutionalized in Britain, and with a royal seal of approval. The Prince of Wales, heir to the British throne, lies behind a move by the University of Wales to establish a professorial chair in the subject, supported from the estate of the late Arthur Koestler, who committed suicide with his wife earlier this year (see *Nature* 302, 93-94; 1983).

Koestler and his wife left £500,000 to support research in parapsychology, and the Koestler Foundation is willing to support a new professorial chair with the interest from a further similar sum. After the bequest became public, the Prince of Wales, who is chancellor of the University of Wales, wrote in his own hand to Dr



C.W.L. Bevan, principal of University College, Cardiff, saying that he would like to see the university make an application for money to Koestler's executors. The idea was unanimously endorsed by the college and by the academic council of the university, and a proposal has now been submitted.

But Wales faces competition from the City University in London and the University of Edinburgh. Professor A. Ellison of City University, president of the Society for Psychical Research and head of the university's electrical engineering department, would like to see a national centre for research in parapsychology based in London, where paranormal phenomena could be subjected to "detailed objective study." He says that such phenomena have often been demonstrated, but that they are not yet repeatable because the parameters for the effects are not known — "it's like electricity in the days before Faraday". Many departments in the university would be willing to support the proposed new professor with their expertise, and Ellison proposes a multidisciplinary advisory council. He would also like to see professional magicians employed to eliminate trickery by subjects claiming paranormal powers. The proposal is to be considered this week by the university's council.

Tim Beardsley

Franklin Valley defended

SIR — On 29 September (p.354), *Nature* published a letter by S.J. Paterson, chief geologist of the Tasmanian Hydro-Electric Commission (HEC) which reveals a lack of understanding of archaeological evidence and interpretation, which issued a threat to the funds available for archaeologists and which questioned the objectivity of archaeologists opposed to the Franklin dam, among whom I was mentioned by name.

I first published an evaluation of Tasmanian prehistory in 1961, and wrote in defence of the Franklin region in 1981 when I learned of its Ice Age habitation sites and noted the meagre and dismissive six sentences on prehistory in the HEC 1979 Draft Environmental Report.

Dr Paterson deplores that the Australian Association of Consulting Archaeologists advised its members to refuse to undertake salvage archaeology, thereby depriving HEC of assistance. He omits that this advice was given as recently as February. From the first archaeological discoveries early in 1981, HEC had two years in which to employ archaeologists.

Dr Paterson also neglected to explain that this advice, in a letter from the association's president to Premier Gray, was given after the then Federal Minister for Home Affairs and Environment had offered massive support for salvage archaeology but on lines that violated both archaeological ethics and the ICOMOS international charter on the protection of monuments and sites, of which Australia is a signatory. The letter also stated succinctly the reasons that led me to oppose dam construction — HEC's inadequate assessment of the environmental impact, its failure to reconsider the dam project even after the recognition of the archaeological importance of the site and its decision to push ahead with roads and construction camps regardless of possible archaeological damage.

The purpose of the frantic field survey conducted while the High Court case was being prepared was not simply archaeological, and does not justify the destruction of one area by the discovery of another. Dr Paterson quotes an impressive number of caves discovered, but experience at Franklin, where only two major archaeological sites and several minor ones were located in some 100 caves and shelters, suggests that other valleys are likely to contain relatively few major archaeological places. Ironically, if the funds spent on this helicopter-assisted but archaeologically leaderless survey had been spent during previous years, the entire region could have been evaluated.

It is to be hoped that assessment will now be carried out, once HEC observes Tasmanian law and provides full details of all discoveries to the Tasmanian Parks and Wildlife Service. Even if these valleys do

contain numerous sites, however, this will merely emphasize the region's truly unique quality. Such a complex of caves preserving Ice Age evidence is unsurpassed anywhere.

Dr Paterson claims to refute my assertion that the Franklin caves may compare in importance with those in south-west France, but he evidently believes that these latter caves are significant only because of their painted walls. However, deep archaeological deposits in the Dordogne have been reexcavated for over a century and provided the classic type sites for most Upper Palaeolithic cultures. They have been the yardstick by which cultural origins have been judged. The remarkable fact is that Tasmanian evidence challenges this Eurocentric version of prehistory. Tasmania should provide evidence for the reconstruction of technological, economic and social life during those same millennia when modern people colonized the globe.

In human terms, these remote places possess remarkable qualities. When people first reached the Franklin Valley, the world was in the grip of its last deep freeze. Those hardy pioneers in this then-open country were the most southerly colonists on the Earth's surface. The limestone and dampness have combined to preserve evidence in these caves in a remarkable manner. Less than a cubic metre of deposit in Kutikina cave produced more stone artefacts and animal bones than the total of finds in all other excavated Tasmanian Ice Age sites, including Cave Bay cave and two in the Florentine Valley.

These sites are also unique in their context. Here are entire river systems occupied in the Ice Age and the sites left intact since then. What was the form of the original colonization? Were all caves occupied simultaneously, or seasonally for specialized functions? Were some sites living places and others spiritual abodes? Dr Paterson considers such caves "merely record occupation by hunting parties" but this undervalues the insight they may give into early society.

The international reality is that south-west Tasmania fully satisfies the criteria for World Heritage listing. Together with two other Australian World Heritage designations, the Kakadu and Willandra regions, there is an opportunity for studying in the Southern Hemisphere complementary examples of adaptation to monsoonal tropics, semi-arid interior and cool temperate forest habitats.

The challenge facing Tasmania is to stop ignoring its past and to develop land management policies that encourage respect for environmental and cultural values.

D.J. MULVANEY

*Department of Prehistory,
Australian National University,
Canberra, ACT 2600, Australia*

India's science base

SIR — In his letter on "Indian expatriates" (*Nature* 3 November, p.10), Dr Malviya makes generalizations ("scientists in India are not terribly science-minded") which imply that his opinions are widely shared. Any scientist, Indian or otherwise, with such attitudes must face the hostility of the Indian scientific community. It is well known that more than half of the Indian expatriates stay on abroad not because they are "terribly science-minded" (mediocre Indian scientists have been employed and offered tenure abroad) but because of the money and the "good life" — which is perfectly understandable.

Indians who oppose the return of their colleagues are not afraid they will have to re-evaluate their scientific research. They may merely be protesting at the deal handed out to them. They have chosen to stay and conduct research, and their efforts should not go unrecognized. If the Indian Government is prepared to invest money, why does it not do so to improve existing facilities and working conditions rather than to build a "technology city"? Why should the attractive incentive be offered only to the expatriate?

Those Indian scientists who have chosen to stay in India even when they have had opportunities to work abroad deserve appreciation for trying to carry on research with the meagre funds that trickle into science in India. Are they to blame?

I also disagree with Dr Malviya's rather pompous declaration that their research is of little importance internationally. A recent survey in *Current Contents* (26, 33; 15 August 1983) shows that India is the most cited of Third World countries and it ranks eighth in the world for the number of articles contributed by its researchers in the international journals sampled. Many Indian scientists try to carry out research in applied fields, for which there is more money available and which has more applications in a developing country like India than does basic research. In addition, Indian scientists make many contributions to journals of theoretical physics and biology.

Dr Malviya's letter seems to me to be little more than the self-justification of a mildly guilt-ridden person. Being an Indian scientist abroad hardly gives one the authority to air views on one's Indian counterparts who have chosen to stay and work against overwhelming odds. It is considerably easier to comment on the need for a technology city in India if one is on the obviously greener side of the fence. The decision to create a technology city may not be the best, but certainly not for the reasons put forward by Dr Malviya.

AMEETA KELEKAR

*St Louis University,
School of Medicine,
St Louis, Missouri 63103, USA*

Fuzzy sets make fuzzy logic

Artificial intelligence has become all the rage, often by government sponsorship. The concept of fuzziness, while interesting, points to few distinctive benefits. Understanding remains the need.

It is a common and legitimate complaint about computers that they are maddeningly literal. One who keeps a list of references to the literature on a computer and who enters one author under the name of JONES cannot afterwards hope to retrieve it by calling JOSNE. And naturally, if JONES is mistakenly entered as JOSNE, and the perpetrator of this simple misspelling does not realize what has happened, the entry may never be retrieved. This puts computers in a poor light compared with people, who can usually tell at a glance when one entry in a printed document is a misprint for some other.

What can be done to remedy these infuriating deficiencies? Formally, at least, there is very little difficulty. It would now be an undergraduate exercise in some computer science course to follow each vain search for an alphabetical data base by the construction of all anagrams (meaningful or not) of the object word and to institute a search for them. The only cost will be measured in computer-time, no great obstacle these days. The drawback is that such a solution, or even the refinement in which (in principle) some empirical knowledge of which mistypings are most common is used to construct a hierarchy of anagrams, is exceedingly inelegant. And the solution of the problem of finding a misspelling of an unknown index-word would require the construction of all anagrams of all possible index-words or otherwise a wooden search through every cell of the computer memory.

No wonder that the search is on for more effective solutions of this kind of problem. The banner under which the search is undertaken is by now familiar as "artificial intelligence", AI for short. Components of it are to be found in publicly sponsored development programmes such as the British Government's response to the report of the Alvey committee on information technology, the collaborative programme of industrial research in Japan known (wrongly) as the Fifth-Generation Computer Programme, the European Esprit programme and elsewhere. In passing, in Britain, it is curious that AI should so quietly have become respectable a decade after a report by Sir James Lighthill assured the Science Research Council (as it then was) that it had better not touch the topic with a barge-pole.

The objective is to embody in machines some of the attributes of human perception and subsequent analysis. So much is un-

controversial. One thing that might be agreed in advance is that solutions of the problem of accurate perception in which the deficiencies of existing machinery are made good by extra hidden programming, the construction of all the anagrams of JONES for example, will not count. The justification for that disdain is that they are bound to be particular solutions of particular problems.

A small step forward would be to replace the search for JONES by a search for some four-letter word derived from the original by the omission of one character. Many would argue that this is in reality a simulation of what people do when scanning a list of index-entries — much of the information on the printed page is redundant, rarely is every letter essential. The snag is that such a procedure might well recover the name JOHNS as well as JONES, but that would rarely be a serious difficulty.

Such a way of tackling the problem of AI seems to have inspired the now fashionable concern with what is called fuzziness, the attempt to build imprecision into programs and even into the machines which such programs instruct. Plainly there are many practical benefits in such developments. Computerized control systems whose input is blessed with spurious accuracy, for example, are capable of hunting (in the technical sense) for an ideal that does not exist; built-in imprecision, perhaps by means of running time averages, is well known to be advantageous in such circumstances. What is much less clear is whether the great flowering of the concept of fuzziness is likely to be as rewarding.

At least one recent book (*Advances in Fuzzy Sets, Possibility Theory and Applications*, Paul P. Wang (ed.), Plenum, New York, \$55) suggests that the movement may have gone too far, at least in its claims to have important practical applications. The notion of a fuzzy set, the invention of L.A. Zadeh of the University of California in 1965, is an entirely uncontroversial extension of respectable mathematical concepts. The conventional definition of a set of points on the interval of some line presupposes that the points themselves are well defined, if not numerically but then by some rule. It is perfectly proper, and interesting, to ask how the properties of sets must be modified if imprecision is built in; plainly quantities representing measure (distance) which may be conventionally additive may lose that convenient property when the sets are fuzzy

(but that may be just the property that AI requires).

What remains unclear is whether the concept of fuzziness as thus defined is capable of being applied in computer problems in ways that are inherently different from much more homespun concepts, for example that the positions of all points or the values of all numbers may be uncertain but nevertheless related to a "true" value by some probability distribution. The devotees of fuzzy theory who contribute to Wang's volume (one third of whom are from the People's Republic of China) insist that their new craft is distinct from probability theory, in practice as well as in principle. The crucial test will be whether they can write computer programs that are markedly more intelligent than can be devised by embedding all numbers in probability distributions, at this level the analogue of the search for JONES which takes account of all anagrams of the word.

While this issue is undecided, it is perhaps fortunate that the prospects for true AI are not wholly bound up in fuzzy sets, and the fuzzy logic by which they must be manipulated. Genuinely parallel computing networks, perhaps operating out of synchrony with each other, offer scope for powerful experiments. There are also opportunities for programmers following some of the interesting demonstrations of the past few years. By now, for example, it is clearly possible to provide programs with such a detailed model of how the human body works, and with such a wealth of physiological data, that the machines which carry them can match the performance of human physicians at the diagnosis of diseases drawn from a restricted set. In several working systems of this kind, the machines do follow human physicians in forming hypotheses and checking them against the data they have been given, making judgements along the way about the significance of different elements of the data, sometimes conflicting. As yet, however, nobody seems to have claimed that machine physicians can better the performance of the human kind.

In the circumstances, it is forgivable that many people regard AI as a field of research, not of development. The task of understanding and then simulating the visual system shows what conceptual difficulties lie ahead. Not the least of these is semantic. When will a machine be able to say "I see JONES!"?

John Maddox

Toxicology

Nitrosamines and human cancer: proof of an association?

from Valda M. Craddock

SINCE the suggestion twenty years ago that nitrite in tobacco smoke might react with the nor nicotine to form N-nitroso-nor-nicotine¹, a compound known to produce cancers of the respiratory tract in rodents, considerable effort has gone into determining nitrosamine levels in the environment, assessing human exposure, and attempting to correlate exposure with cancer incidence. The results have been sufficiently encouraging to compel people to continue, but until recently they have been far from conclusive. Now, as a recent meeting on N-nitroso compounds² demonstrated, many agree with K. Preussmann (German Cancer Center), that proof of an association with human cancer has been found.

Convincing epidemiological evidence shows an association of oral cancer in the southern states of America with snuff dipping, as distinct from snuff sniffing or tobacco chewing (D. Winn, National Cancer Institute). Most telling is the fact that cancers develop in the mouth at the site where the tobacco was retained. While several carcinogens are known to be present in chewed and smoked tobacco, nitrosamines are the only carcinogens to have been detected in snuff, the levels exceeding by two orders of magnitude those in other consumer products (D. Hoffmann, Naylor Dana Institute, Valhalla). This seems to be as near as one is likely to get to proof from epidemiological evidence for an association of human cancer with a specific group of chemicals. The work also shows that nitrosamine levels found in the environment can be sufficient to cause cancer. Proof of the association in this one case makes it more likely that a link will be proven in other cases where nitrosamine involvement is already suspect.

Careful studies in regions with dramatic differences in incidence of oesophageal cancer in China, South Africa, and Iran, and in areas with a high incidence of stomach cancer in South America, have not, however, revealed any simple correlation between exposure and cancer incidence, owing perhaps to the many factors involved in the etiology of the disease. One complication is that total exposure to nitroso compounds is not determined by the amount ingested, as nitrosation can occur within the body. Whenever amines, amides or ureas come into contact with nitrite, either in the mouth or in the stomach, nitroso compounds may be formed. The extent of nitrosation depends on the amines ingested, on the presence of catalysts and inhibitors such as vitamins A and C and phenolics in the diet (M. Archer,

Ontario Cancer Institute, Toronto; H. Stich, Environmental Carcinogenesis Unit, Vancouver), on the ingestion of nitrate, and on the presence of bacteria which can reduce nitrate to nitrite. The potential of individuals for nitrosation therefore varies, and needs to be known before exposure to nitroso compounds can be assessed.

Several groups have tackled this problem by determining urinary nitroso-proline after ingestion of proline and nitrate (S. Lu, Cancer Institute, Beijing; H. Bartsch, IACR, Lyon; P. Reed, Hammersmith Hospital, London; J. Elder, University of Manchester; D. Wagner, MIT). Interesting results have included the discovery in urine of a previously unknown amino acid, N-nitroso-thiazolidine-4-carboxylic acid. No simple correlation between exposure to total nitrosamines — dietary plus those formed *in vivo* — and cancer incidence was revealed.

It is possible that nitrosation of proline is not an ideal test, as different precursors vary in their pH optima for nitrosation. Achlorhydria and a variable gastric pH are associated with gastric cancer. It is therefore necessary to characterize the compounds formed in human stomach, and to test them individually (H. Bartsch).

Another factor that was previously underestimated is the effect of diet on the susceptibility of host tissues. The levels of carcinogens to which humans are exposed are generally low compared with those used in animal experiments; but in experiments in which low doses of nitrosamines have been used, modifying factors have been shown to have a dramatic effect. Thus zinc deficiency can increase the incidence of methyl-benzyl-nitrosamine-induced oesophageal cancer from 15 per cent to 80 per cent². Low levels of exposure may be sufficient to cause cancer if the oesophagus is predisposed, so that dietary deficiencies or excesses may play a predominating role in determining whether an individual develops cancer or not. Given the vitamin and mineral deficiencies in the diet of the Chinese in areas with a high incidence of oesophageal cancer (S. Lu), nitrosamines may well be the initiators of the disease, even where exposure, considered alone, does not correlate with cancer incidence. Again in stomach cancer, the high salt, micronutrient deficient diet which damages the mucosa and leads to achlorhydria and possibly to increased formation of nitrosamines also causes regenerative hyperplasia, a factor known to predispose various organs including stomach to carcinogens (S. Tannenbaum, MIT).

To unravel the multifactorial etiology of the disease, an understanding of the molecular mechanisms by which nitroso compounds act is essential. It seems likely that, at least with low molecular weight compounds, an O⁶-alkylguanine-DNA adduct is the relevant lesion, and that replication of DNA must take place before the lesion has been removed by cellular repair processes (M. Rajewsky, University of Essen; P. Kleihues, University of Freiburg; A. Pegg, Pennsylvania State University; V. Craddock, MRC Toxicology Unit). Obviously there are many ways in which environmental factors could promote or inhibit the sequence of events leading to neoplasia: formation of nitroso compounds *in vivo*, metabolic activation where necessary, reaction with DNA, and replication of DNA before repair has occurred.

An important feature of the meeting was the long-awaited result of an animal experiment in which 5,120 rodents were treated with 15 dose levels of various nitrosamines for life (R. Gray, Radcliffe Infirmary, Oxford). Although the data from this colossal undertaking have yet to be fully evaluated, effects were 'clearly measurable' at doses of 0.01 mg per kg per day, an order of magnitude lower than previously described. While the rats used in these experiments had a life span of 3½ years, man may be exposed for 70 years or more. There is also the strong possibility of the additive or synergistic effects of other carcinogens, and of the action of potentiating factors in the human situation. The idea of determining a 'safe, no-risk' level of exposure is obviously futile. To pronounce a ban on nitrosamines in the environment would be as absurd as the idea of banning various other ubiquitous carcinogens, such as sunlight or polycyclic hydrocarbons. Also it is difficult to see how nitrosation of tobacco-specific alkaloids can be prevented. Detection of cancer-prone individuals may be a more feasible proposition (R. Setlow, Brookhaven National Laboratory), but the value of giving additional warnings to those who do not want to know is dubious. The plan to persuade governments to subsidize farmers to grow alternative crops is the most hopeful solution to the tobacco problem. However, there are many ways in which nitrosamine levels in food and in manufactured products can be lowered, *in vivo* nitrosation can be reduced, and potentiating risk factors, such as micro-nutrient deficiencies, can be remedied. These measures should reduce the incidence of diet-related and environmental cancers. Hopefully we are 'now on the threshold of a period that may be dominated by preventative intervention'³. □

Valda M. Craddock is in the MRC Toxicology Unit, Carshalton, Surrey SM5 4EF

1. Druckrey, H. & Preussmann, R. *Naturwissenschaften* 49, 498 (1962).
2. Fong, L.Y.Y., Sivak, A. & Newberne, P.N., *J. natn. Cancer Inst.* 61, 145 (1978).
3. van Rensburg, S.J. *S. African Cancer Bull* 26, 153 (1982).

Nitrogen fixation symposium

Molecular genetics: out of the laboratory into the field

from Allan Downie

THE Fifth International Symposium on Nitrogen Fixation, an interdisciplinary meeting which ranged from the co-ordination chemistry of nitrogenous compounds to the ecology of diazotrophs, was one of a series aimed at bringing together scientists with interests from all areas relating to nitrogen fixation. It is clear that molecular biology techniques are now having a significant impact on the agriculturally important process of biological nitrogen fixation.

The fundamental research carried out on the molecular genetics of nitrogen fixation in *Klebsiella pneumoniae* has been exploited by those analysing nitrogen fixation in bacteria less amenable to genetic manipulation, including, for example, the cyanobacterium *Anabaena* and *Rhizobium*, both of which participate in symbioses important in agriculture.

Several genes involved in nitrogen fixation in *Anabaena* were cloned using DNA homology by Haselkorn *et al.* (University of Chicago). The glutamine synthetase gene, important in the assimilation of fixed nitrogen, appears to have two promoters, one of which is specifically switched on when *Anabaena* is actively fixing nitrogen. Since nitrogen fixation in *Anabaena* occurs within differentiated cells (heterocysts) it is possible that genes involved in nitrogen fixation are regulated in a similar way to genes involved in spore formation in *Bacillus* and it was suggested that there may be σ -factors which determine the specificity of the RNA polymerase for different promoter sequences. The three structural genes for nitrogenase have also been cloned from *Anabaena* and shown to have a different arrangement from those in *K. pneumoniae*, with the *nifH* and *D* genes adjacent to each other but separate from *nifK*. For reasons that are as yet obscure, *Anabaena* appears to contain a reiteration of (part of) the *nifH* sequence.

The theme of repeated sequences was one of particular interest to those working with *Rhizobium*. Quinto *et al.* (*Nature* 299, 724; 1982) were the first to report the reiteration of structural nitrogenase genes in *R. phaseoli*. At this meeting the genes were also shown to be present in *R. trifolii* (Shine *et al.*, Australian National University, Canberra) and *R. meliloti* (Better *et al.*, University of California, San Diego). In *R. meliloti* a series of six highly conserved DNA sequences was reported and these appeared to represent transcriptional control

regions for genes expressed by *Rhizobium* during symbiotic nitrogen fixation. Nucleotide sequencing revealed that promoter-like sequences were highly conserved and, surprisingly, that there was more than 80 per cent homology over a 160bp region spanning the sites of initiation of transcription. Downstream of one of these promoters an open reading frame was identified corresponding to a second *nifH*-like gene. Surprisingly the similarity in sequence ended abruptly and the function (if any) of this partial gene remains to be established.

The nodulation genes from *R. meliloti* (Long *et al.*, Stanford University) *R. leguminosarum* (Downie *et al.*, University of East Anglia) *R. trifolii* (Shine *et al.*, Scott *et al.*, DSIR, New Zealand) and *R. parasponium* (Marvel *et al.*, Harvard University) have been cloned and shown to transfer the plant-host-range determinants between different species of *Rhizobium* and even between *Rhizobium* and *Agrobacterium*. Progress in this area of molecular biology is rapid and a detailed knowledge of the genes and their sequences may soon be available; the next step will be to analyse their biological role. One or more of these genes will almost certainly be important the initial stages of the plant-microbe interaction.

The role of plant lectins and bacterial polysaccharides at the early stages of the plant-microbe interaction was an area of lively discussion. Bhuvaneswari *et al.* (C.F. Kettering Research Lab., Yellow Springs) and Solheim (University of Tromsø) reported that *R. trifolii* when cultured in the presence of white clover produced a factor which induced branching of the clover root hairs. Thus the mature root hairs (which are not usually susceptible to infection by *Rhizobium*) could be induced to branch, forming a new growing tip which was susceptible to infection. A (partially) purified preparation of the 'branching factor' (consisting mainly of oligosaccharides) was shown to enhance nodulation, particularly in the mature region of the roots where less nodulation usually occurs. Stacey (University of Tennessee) identified a plant-produced component which has a role in the early steps of the *R. japonicum*/soybean symbiosis. Using a mutant of *R. japonicum* which formed few and delayed nodules on soybeans, he found that a component on soybean root exudate could partially reverse the mutant phenotype of the *Rhizobium* strain. The plant-encoded component appeared to be a protein with properties similar to those of a lectin. Two

of the steps in the communication between bacterium and plant thus appear to have been outlined.

There has been a growing interest in the molecular biology of plant genes involved in nodulation and nitrogen fixation. Bisseling *et al.* (Agricultural University, Wageningen) and Verma *et al.* (McGill University, Montreal) reported the characterization of several nodule-specific, plant-encoded RNA species which code for nodule proteins (nodulins). Analysis of one of these components by Verma *et al.* showed that it was uricase, an enzyme important in the assimilation of fixed nitrogen. Ultrastructural analysis of nodules using labelled antibody to uricase showed that it was located not in the plant cells which contained bacteroids, but in adjacent uninfected cells. Using a different approach (subfractionating nodules) Schubert *et al.* (Monsanto Agricultural Products, St Louis) also identified a number of enzymes important in nitrogen assimilation in uninfected plant cells.

The identification of the locations of plant-encoded nodule enzymes is essential for an understanding of symbiotic nitrogen fixation. The location of the oxygen-binding protein leghaemoglobin has been a source of controversy although evidence has been accumulating that it is present within the plant cell cytoplasm rather than within the peribacteroid space (although the definitive experiments have yet to be done). As discussed by Robertson (DSIR, New Zealand) and Appleby (CSIRO, Canberra) the presence of leghaemoglobin in the cytoplasm suggests that it does more than supply oxygen to the bacteroids. Because the cytoplasmic concentration of free oxygen would be very low (in the μM range), the question arises as to whether the mitochondria (and other oxygenases) could operate satisfactorily at low oxygen tension. The future analysis of the plant's role in the symbiosis looks to be a promising area, especially now that Jacobsen (University of Groningen) has isolated mutant strains of peas which can be nodulated at levels of fixed nitrogen that would normally inhibit nodulation, and Larue *et al.* (Cornell University) have isolated a nodulation deficient mutant of pea.

Many other areas of the conference were of great interest. For example Haaker (Agricultural University, Wageningen) questioned the generally accepted mechanism of electron transfer within the nitrogenase enzyme complex; Leigh (University of Sussex) questioned if dinitrogen was bound end-on or side-on at the active site of nitrogenase; and Witty *et al.* (Rothamstead Experimental Station) questioned the validity of the use of the acetylene reduction assay which is usually used to assess the levels of nitrogen fixation occurring in plants. □

Allan Downie is in the Department of Genetics, John Innes Institute, Colney Lane, Norwich, NR4 7UH.

The Fifth International Symposium on Nitrogen Fixation was held from August 28-September 3, 1983 at Noordwijkerhout, The Netherlands. The proceedings will be published by Nijhoff, The Hague, in January 1984.

Atmospheric physics

Natural lasers on Venus and Mars

from F. W. Taylor

THE study of non-equilibrium processes in planetary atmospheres is in vogue at present as interest grows in the stability of our natural environment. It has long been realised that the Earth's upper atmosphere cannot be in thermodynamic equilibrium, and during the last decade astronomers have made telescopic observations¹ of non-equilibrium processes taking place in the upper atmospheres of our Earth-like neighbours, Mars and Venus. A preliminary analysis by Michael Mumma of Goddard Space Flight Centre in Maryland, and his colleagues², indicated that processes analogous to optical pumping in laboratory lasers were taking place and led to the coining of the term 'natural lasers'. Now, Deming and colleagues from the Goddard team have taken new observations³ of emission from Mars and Venus at wavelengths near 10 micrometers, and modelled⁴ them to show that stimulated emission — the effect which makes lasers so powerful — accounts for up to seven per cent of the total emission. This is not a large amplification factor by laboratory standards, particularly for a CO₂ laser. But, the authors speculate that the sheer size of the natural lasers could make them useful tools in the future for communicating with distant civilizations far beyond our own planetary system.

Such a far-reaching application is at least conceivable because of the collisions which are constantly taking place between the molecules of the air and infrared radiation. When a molecule absorbs a photon, it stores its energy by making the transition to a higher vibrational or rotational energy state. What happens next involves a web of processes, including natural relaxation to the original state (with re-emission of the photons), relaxation or excitation to a different state (with emission or absorption of a photon of different energy and therefore wavelength), or collisional de-excitation. In the latter case the molecule relaxes by colliding with another and imparting additional kinetic energy to both. In other words, the gas absorbs the radiation and gets warmer, a familiar enough occurrence.

At low pressures, however, such as are found in upper atmospheres where the air is very rarefied (around one millionth of normal surface pressure on Earth for the regions of natural laser emission on Venus and Mars) collisions are relatively rare and the 'thermalization' of photons is relatively infrequent. The molecules thus tend to shuttle the energy from absorbed photons around their myriad energy levels in a very

complicated fashion. The process is very difficult to model but is important, not only as basic physics, but because it controls the radiative energy balance of the upper atmosphere and hence its structure and dynamics. Carbon dioxide is the most important molecule in this context and, its behaviour is strongly modified by the availability of collision partners of different types.

Venus and Mars have atmospheres of nearly pure CO₂ and Deming *et al.* believe that the molecules are excited by the absorption of energetic photons from the Sun in the same way as a laboratory laser is pumped by an electrical discharge. Thermally emitted photons from the lower atmosphere on their way to space then can collide with the excited molecules and stimulate the emission of a second identical photon. Since this creates two photons where before there was one, light amplification by the simulated emission of radiation has taken place — laser action. A

complication to this simple picture is the fact that most solar photons are so energetic that the absorbing molecules are too highly excited to emit the necessary low-energy ten micro-metre radiation. Deming and Mumma⁴ postulate that partial de-excitation takes place by collisions between molecules until energy at the right level can be emitted. They back up their theory with detailed (but still incomplete, because of the state of the art) model calculations which show it to be not only plausible but capable of accounting rather closely for their observations.

Deming and Mumma end with speculations on how the natural laser on Mars might be turned into a man-made one of enormous power by placing large mirrors in orbit, parallel to each other. By passing the photons back and forth, the intrinsic gain of a few percent could be multiplied and an intense beam of infrared radiation, presumably suitably modulated with messages of good will, directed towards neighbouring stellar systems. □

F. W. Taylor is in the Department of Atmospheric Physics, Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU.

1. Johnson, M.A., Betz, A.L., McClaren, R.A., Sutton, E.C. & Townes, C.H., *Astrophys. J.* **208**, L145–L148 (1976).
2. Mumma, M.J. *et al. Science* **212**, 45 (1981).
3. Deming, D. *et al. Icarus* **55**, 347 (1983).
4. Deming, D. & Mumma, M.J., *Icarus* **55**, 356 (1983).

Ciba Foundation symposium

Human cataract formation

from George Duncan

CATARACT is by far the most common cause of blindness throughout the world. As a result of research efforts on a very broad front, which have been reported at a recent symposium*, we are closer both to understanding the molecular mechanisms involved and to developing preventive drug therapy for certain types of cataract.

Progress in understanding cataract formation has been impeded by our lack of understanding of the physical basis for lens transparency: if the proteins in the lens were simply distributed at random they would make the lens opaque. It appears that a paracrystalline structure is not required for transparency as had previously been believed, but simply a degree of relatively short-range order in the macromolecules making up the lens (see *News and Views* **302**, 383; 1983). Breakdown of this order leads on the macroscopic scale to phase separation with accompanying light scattering so that the lens becomes opaque. Several groups are taking advantage of the natural optical accessibility of the lens and are developing *in vivo* techniques for studying the

mechanisms underlying this disorder. Preliminary experiments with a cataract model system examined by X-ray scattering in the rabbit indicate, for example, that changes in protein diffusivity can be detected *in vivo* before opacification occurs. (G. Benedek, Massachusetts Institute of Technology).

It is important to realise that the entire lens is not always involved in the cataract process. Much effort is being devoted to identifying and cataloguing the specific areas affected in different types of cataract (L.T. Chylack, Harvard Medical School). Human lens microdissection techniques, both *in vitro* and, somewhat spectacularly, *in vivo*, are now far advanced (J. Horwitz, Jules Stein Eye Institute, New York) and it is possible to detect specific changes in lens proteins from opaque regions that are not present in neighbouring clear regions. Differences between clear and opaque regions can also be discerned in human lenses using calcium ion-sensitive microelectrodes. A 10 to 100-fold increase in free calcium occurs in regions with highly localized opacities, while the surrounding clear regions remain at the normal level of less than 10 μ M calcium. The rise in free calcium may also be responsible for isolating the affected area, since gap

*The symposium on Human Cataract Formation (Ciba Foundation Symposium 106) was held in London, 25–27 October 1983; the proceedings will be published by Pitman, London, in summer 1984.

junctions in lens fibre membranes appear to be closed when exposed to increased calcium levels in the lens (G. Duncan & T. Jacob, University of East Anglia).

Calcium is also involved in stabilizing the lens cytoskeleton. The actin network, which is widely distributed in the lens, interacts with elements, spectrin and calmodulin (H. Maisel, Wayne State University). Vimentin, an unusual protein to find in cells of ectodermal origin, is located in the lens epithelium and superficial fibres and appears to interact mainly with membrane gap junctions, while actin reacts mainly with the plasma membrane (H. Bloemendal, University of Nijmegen). In terms of gap junction physiology, it is interesting that the lens protease which cleaves vimentin is, in fact, calcium-activated.

Certain lens crystallin subunits also interact strongly with the membrane. A possible reason for this has been suggested following important advances in our knowledge of lens protein structure, gained from X-ray diffraction (T. Blundell and C. Slingsby, Birkbeck College) and sequencing analyses (H. Bloemendal, University of Nijmegen; J. Piatigorsky, National Eye Institute, Bethesda). Having established the repeating-motif structure of γ -crystallin (four structural motifs in two domains), it has been possible to predict the structure of certain β -crystallin subunits from their amino acid sequence in view of the homology between the β - and γ -crystallins. Some β -crystallin subunits have hydrophobic extensions — these arms may be responsible for anchoring the protein to the membrane.

Rapid advances in our knowledge of crystallin gene structures were also reported. The exons encoding the four structural motifs of the β -crystallins are separated by introns (J. Piatigorsky). In the γ -crystallin genes, which form a clustered family, the exons encoding the two structural motifs in each domain are fused, however, those encoding each domain of the protein are separated by an intron (J.G.G. Schoenmakers, University of Nijmegen). Analysis of the sequence and secondary structure of mRNAs for δ - and β -crystallin polypeptides raised the possibility that there may be two functional initiation sites for translation. If this novel situation is true, it makes it possible that ion-induced changes in the ratio of synthesis of crystallin polypeptides may be related to mRNA secondary structure. This is of considerable interest to the cataract research field, where changes in lenticular ion concentrations have been linked with lens opacity and alterations in crystallin synthesis.

The stability of the β and γ crystallins is critically dependent on the maintenance of several sulphhydryl groups in the reduced state. It is therefore not surprising to find that the lens contains the highest tissue concentration of glutathione. It is also not surprising that oxidation is an early event in

cataractogenesis. Although oxidation was first invoked to explain the changes in protein structure that occur in nuclear cataract (J.J. Harding, University of Oxford), loss of glutathione and an increase in fibre membrane peroxidation products also occur in cortical cataract (V. Reddy, Oakland University). Significant levels of oxidants, including hydrogen peroxide, are in fact found in human aqueous humor (A. Spector, Columbia University) and the epithelial cell enzyme glutathione reductase is largely responsible for minimizing internal oxidation (F. Giblin, Oakland University). Oxidation may be a much more general cataract mechanism than had previously been believed since autoxidizing dicarbonyl compounds are produced from glucose metabolism in the lens (M.J.C. Crabbe, University of Oxford). It has long been known that cataract is associated not only with diabetes, but also simply with a raised serum level of glucose (R. Clayton, University of Edinburgh). Cataract in this case was considered to arise solely from the production of impermanent sugar alcohols by the enzyme aldose reductase (P. Kador and

J. Kinoshita, National Eye Institute, Bethesda). It is comforting to know that the drugs developed to inhibit aldose reductase (and now undergoing clinical trial) also inhibit the formation of glucose autoxidation products.

It was pointed out as a postscript to the meeting that, since it dealt with the X-ray diffraction analysis of lens crystallins and the microstructure of the lens cytoplasm, it is fitting that it was held on the 150th anniversary of the publication of a paper (*Phil. Trans. R. Soc. B* 123, 323; 1983) in which optical diffraction techniques were first used in any biological system to obtain quantitative information on the structural components of that tissue (in this case the fibres of the lens). Since the author, Sir D. Brewster of the University of St Andrews, included cell surface invaginations and filamentous cytoplasmic inclusions in his drawings of lens fibres, he would have been quite in his element at this stimulating symposium. □

George Duncan is Reader in Biophysics in the School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ

Deep Sea Drilling Project

Testing the Vail depositional model

from the staff of Leg 95

DURING August and September, 1983, *Glomar Challenger* cored two locations on the New Jersey Transect, providing a geological test of the seismostratigraphic model of depositional sequences proposed by P.R. Vail and his colleagues. The New Jersey Transect was conceived as a means of analysing the stratigraphic framework and depositional history of the North American passive continental margin, beginning with wells on the New Jersey Coastal Plain and terminating at DSDP Site 603 on the lower continental rise. The abundance of available geologic and geophysical data from the coastal plain, shelf and upper slope provides an extensive data base for the shoreward segment of this transect. United States Geological Survey (USGS) multichannel seismic line 25 serves as the key control line, providing detailed seismostratigraphic data across the depocenter of the Baltimore Canyon Trough and extending from the inner shelf to the middle continental rise.

Phase one of the New Jersey Transect was completed by DSDP-IPOD Leg 93, which drilled two sites on the upper rise (Sites 604 and 605) and also cored a 1,576 m section at Site 603, to establish the basinward termination of the transect. Phase two of the New Jersey Transect (Leg 95) established two additional reference sections; one on the middle slope (Site 612; 1,404 m water depth) and one on the upper

rise (Site 613; 2,332 m water depth). Both sites were cored and a suite of downhole geophysical logs was obtained at each.

Before Leg 95 drilling began, C.W. Poag and co-workers established the broad regional stratigraphic and depositional framework of the shelf and upper slope of New Jersey. In particular, they documented a series of depositional episodes punctuated by widespread erosion, whose pattern resembles a response to sea level fluctuations, such as envisioned in the widely discussed Vail model. Coring at Sites 612 and 613 shows that the Cenozoic and upper Cretaceous depositional sequences of the New Jersey slope and upper rise fit very well into the previously established shelf-slope framework.

For example at Site 612, only two (upper Eocene and lower Oligocene) of nine major stratigraphic units are conformably juxtaposed. The remaining seven units (Campanian, Maestrichtian, lower Eocene, middle Eocene, upper Miocene, Pliocene and Pleistocene) constitute separate depositional units whose boundaries are erosional unconformities. These unconformities are widespread on the slope and shelf, and five of them have counterparts in the Vail model. A sixth represents a long hiatus (late Oligocene to late Miocene) that encompasses three of Vail's unconformities.

At Site 613, four major Cenozoic sequence boundaries were penetrated

(lower Eocene/middle Eocene, middle Eocene/upper Miocene, upper Miocene/Pliocene, Pliocene/Pleistocene). The latter three are unconformable contacts with equivalents at Site 612 and on the shelf, but only the last two were recovered. The fourth sequence boundary comprises a disturbed 'zone' of intense slumping, separating lower and middle Eocene strata.

One of the most interesting new developments regarding unconformities along this segment of the margin is our documentation of a long hiatus (25–37 Myr) at both sites between the late Miocene and the early Oligocene or middle Eocene. Extrapolation of the core data along the grid of seismic lines suggests that upper Eocene, lower and upper Oligocene and lower and middle Miocene strata are virtually absent from the lower slope and upper rise. We cannot, however, rule out the possibility that very thin layers of these units might exist undetected by the limited resolution of our seismic lines.

Seismic lines which follow the depositional strike near Site 613 reveal a series of stacked buried channels, whose ages and origins have challenged seismostratigraphers since they were discovered. Under the influence of the Vail model of sea level change, the general consensus has been that the channeling resulted from a mid-Oligocene sea-level fall. Drilling at Site 613 revealed that one series of channels was formed between the early and middle Eocene. The lower Eocene surface is eroded from the shoreline across the shelf through Sites 612 and 613, and basinward.

Some channels completely penetrate the lower Eocene strata, cutting deeply in the underlying Paleocene section. At Site 613, however, which appears to have been near the slope-rise transition during the early and middle Eocene, the presence of frequent slumps indicates a depositional regime in which displaced sediments were accumulating. These base-of-the-slope accumulations increased during the middle Eocene and eventually filled the channels on the lower Eocene-Paleocene surface.

A second series of channels is stacked upon the middle Eocene surface. The middle Eocene unit has had a complex history

of erosion, as revealed by a wide outcrop belt now present along the lower slope, and by truncated reflections in its upper part. It has presumably undergone several periods of erosion since the late Eocene, and is still being worn away along its outcrop. Middle Eocene clasts were recorded in the Miocene and Pleistocene strata at Site 613 and in surficial piston cores along the lower slope and rise.

Filling of these middle Eocene channels took place during the late Miocene, as revealed at Sites 604 and 613, perhaps during the low stand of sea level associated with the Messinian salinity crisis. The coarse sands, gravels, and lithoclastic conglomerates at Sites 604 and 613 indicate that the channel fill came from the shelf and was dumped on the lower slope and upper rise in rather chaotic fashion.

These findings demonstrate that with

continuously cored, shallow-penetration sections, carefully placed on seismic transect lines, we can easily obtain the fundamental geologic data necessary to unravel the complex Cenozoic stratigraphy and depositional history of sediment-rich passive margins. The concept of multi-site transects has developed late in the DSDP program but the immense value of their systematic approach to margin evolution has been amply demonstrated by the results of such legs as 78, 79, 80, 93 and 95. Moreover, we would hope that the sections now drilled constitute only the initial steps toward a more comprehensive appraisal of margin development. The New Jersey Transect, in particular, should stimulate new proposals for additional sites along Line 25 and its joining seismic grid, especially those aimed at deeper targets within the Mesozoic sequences. □

Oncogenic intelligence

More *rasmatazz*

from Peter Newmark

DESPITE the wealth of information that has emerged in the last 18 months about normal human *ras* genes and the oncogenes that arise from them by single point mutations, several vital questions remain unanswered. First, what causes the point mutation? And second, what are the functions of the proteins encoded by the *ras* genes and oncogenes? Papers in this issue suggest one answer to the former question and a new way to ask the latter.

In the article beginning on page 658, M. Barbacid and his colleagues show that a point mutation distinguishes an H-*ras*-1 oncogene of carcinogen-induced rat mammary carcinomas from the normal rat *ras* gene and suggest that the mutation is a direct consequence of the reaction with DNA of the carcinogen. Nine carcinomas which appeared some months after a single injection of 50 day old female rats with nitroso-methylurea, were examined. Each had a similar H-*ras*-1 oncogene. One oncogene that was sequenced, differed from the normal rat H-*ras*-1 gene simply by having deoxyadenosine (A) instead of deoxyguanosine (G) as base number 34. The predicted result is the substitution of glycine by glutamic acid as the twelfth amino acid of the p21 protein encoded by the gene (just as human *ras* oncogenes often have a mutation that leads to the replacement of their glycine at the equivalent position in p21).

Barbacid *et al.* point out that nitroso-methylurea is an alkylating agent that has a preference for methylating Gs. Once methylated, a G might base pair with thymidine instead of the normal deoxycytosine; subsequently the methylated G would be replaced by A. Hence G would have been substituted by A. As the authors point out, they have only proved the substi-

tution of G by A in one carcinoma but if that is invariably the case it will strongly support their suggestion for the generation of the point mutation. In any case their animal model promises to be very useful in the study of *ras* genes and oncogenes.

So, too, does yeast now that two laboratories have shown it to contain genes closely related to mammalian *ras* genes. Gallwitz *et al.* (p. 704) came across their *ras* gene by chance, at first not recognizing the identity of the sequence close to a yeast actin gene. By contrast, DeFeo-Jones *et al.* (p. 707) deliberately set out to look for *ras*-like genes in yeast. They found two and sequenced one. Comparing its predicted protein with that from the human *ras* gene, the same amino acids are found in about 60 per cent of the positions; the equivalent figure from Gallwitz *et al.* is about 40 per cent. (Incidentally, the Gallwitz *et al.* protein would have glycine as amino acid 12, as do normal human and rat *ras* genes, whereas the DeFeo-Jones *et al.* protein would not, as in many *ras* oncogenes.)

Given the variety of clever tricks that can be played with yeast genes (see K. Struhl *Nature* 305, 391), it may be easier to learn something of the function of *ras* genes in yeast than in mammals. One obvious experiment would be to replace the normal yeast *ras* gene with a disrupted version and observe the resulting phenotypic effects on yeast — in particular, to check if they are similar to those induced by disruption of cell cycle genes in yeast. A more subtle variation on this theme would be to find a temperature sensitive mutant of the yeast *ras* gene. Experiments of that kind are already underway in 'yeast' laboratories and the results are eagerly awaited. □

Peter Newmark is Deputy Editor of *Nature*.

The scientific party aboard *Glomar Challenger* for Leg 95 of the Deep Sea Drilling Project, International Phase of Ocean Drilling, consisted of: C. Wylie Poag (U.S. Geological Survey, Woods Hole), A.B. Watts (Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York), Michel Cousin (Université Pierre et Marie Curie, Département de Géotectonique, Paris), David Goldbert (Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York), Malcolm B. Hart (Plymouth Polytechnic, Department of Environmental Sciences, Plymouth), Kenneth G. Miller (Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York), Gregory Mountain (Lamont-Doherty Geological Observatory, Columbia University, Palisades, New York), Yuji Nakamura (University of Tokyo, Laboratory for Earthquake Chemistry, Nakano, Tokyo), Amanda Palmer (Princeton University, Department of Geological and Geophysical Sciences, Princeton), Paul A. Schiffelbein (Scripps Institution of Oceanography, La Jolla), B. Charlotte Schreiber (Queens College (SUNY), Department of Earth and Environmental Science, Flushing, New York), Martha Tarafa (Woods Hole Oceanographic Institution, Chemistry Department, Woods Hole), Jean E. Thein (Geologisches Institut der Universität Bonn, Bonn), Page C. Valentine (U.S. Geological Survey, Woods Hole), Roy H. Wilkens (Department of Earth & Planetary Sciences, Massachusetts Institute of Technology).

Developmental biology

A new homeotic gene

from Peter Lawrence

To all old-fashioned biologists, and to most modern ones, it is obvious that animals are not ill-assorted bags of fibroblasts, neurons and other specialised cells — yet we do not have much idea why they are not. One of the biggest challenges in embryology is to understand how cell types arise together in specific patterns. Homeotic genes offer a route to attack this problem because a change in the DNA sequence of one gene can comprehensively transform a single cell, and all its descendants, into parts of another organ (for example, a piece of the leg of a fly into a piece of an antenna). Homeotic genes are often in the news these days and there have been two recent advances that have brought them into the molecular era. First, the DNA for the best known homeotic locus (the bithorax complex) has been purified and analysis has begun¹. Second, it is now possible to hybridize molecular probes to sections of *Drosophila* and to find out when, and in which cells, homeotic genes are being transcribed^{2,3}.

The difficulty with homeotic genes, at least for the biologist looking for general principles, is that they have only been found in insects. There is good news therefore in the paper by Dr. Iva Greenwald and her colleagues which has just appeared in *Cell*⁴ for it reports the discovery of a homeotic gene (unmnemonically called *lin-12*) in the nematode, *C. elegans*. Mutations in this gene are of two types; recessive loss-of-function mutations (*lin-12(o)*) and dominant gain-of-function mutations (*lin-12(d)*) that transform developing cells in related but opposite directions. For example, in the gonad there are a pair of cells whose ancestry does not wholly predict their fate — in a wild type one cannot, by following the steps of cell lineage in an individual, predict which of the pair will become a type *A* cell and which will become a type *B* cell — but there is always one of each in each worm. In *lin-12(o)* they both become *A*, while in *lin-12(d)* they both become *B*. The same principle applies to several other pairs of cells elsewhere in the body. The cells affected by mutations in *lin-12* are usually those whose fate is not completely defined by their ancestry and whose development appears to depend on particular cell interactions (because laser ablation of one such cell, or its neighbours, has effects in addition to the elimination of the ablated cell and its descendants).

The opposing properties of the loss-of-function and gain-of-function mutations suggest that, in the wild type, the amount of gene product is directly engaged in determining the fate of certain cells so that type *A* cells need less product than type *B* cells.

Understanding the wild type role of a gene is more difficult than describing the phenotypes of mutants but Greenwald and colleagues have tried to build a case that the wild type role of the *lin-12* gene is to act as a "binary switch to control decisions between alternative cell fates". Apart from the suggestive phenotypes of the mutations themselves, the authors have used a temperature sensitive form of *lin-12(d)* to define the time when changes of level in gene product can change the cell fate. They find that this time occurs at about the latest stage in development that cell fate can be altered by killing nearby cells. These observations suggest, but do not establish, that the extra gene product directly affects the cells at the time that their fates are being determined.

lin-12 is clearly a homeotic gene in the loose sense (even in *Drosophila*, homeotic genes are defined loosely) and there are some interesting unanswered questions. In the wild type, does the gene act only in

some cells and not at all in others, or are we seeing in the mutants some threshold effects in the least canalized steps of the cell lineage? If the action of the gene is restricted to a certain set of cells, what other properties define that set? Most important: is the wild-type role of the gene to control the activity of other genes directly or does it just provide a product necessary for the nuts and bolts of cellular diversification? In other words is *lin-12* a gene whose only role is to direct cells along particular developmental pathways (like the elements of the bithorax complex, or the *transformer* genes of the nematode or the fly) or does it have a general function in cell physiology which is critical only to those cells altered by *lin-12(o)*? Whatever the answers to these questions I am sure we will learn more from *lin-12* in the future. □

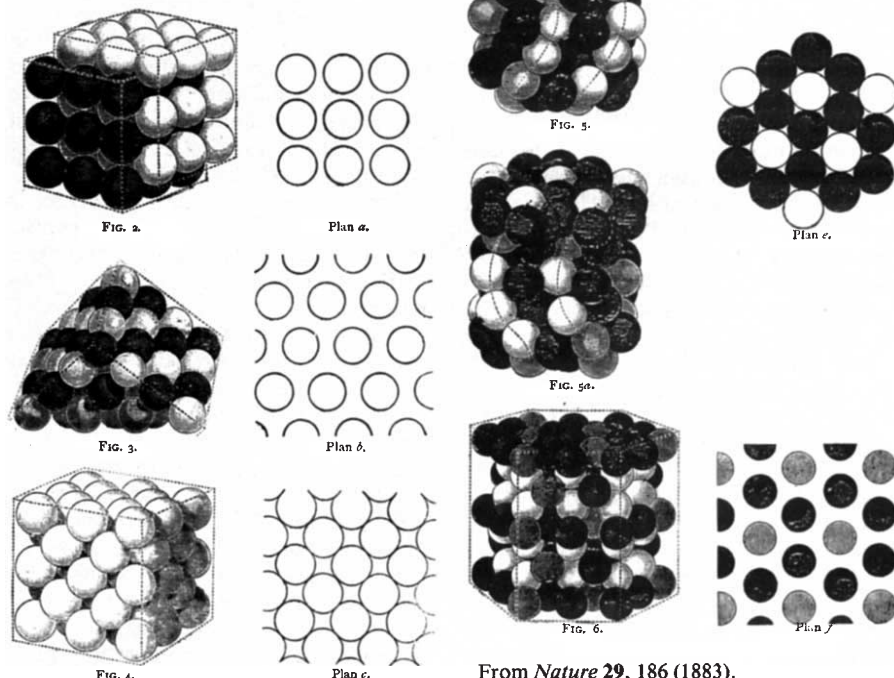
Peter Lawrence is in the MRC Laboratory of Molecular Biology, University Medical School, Hills Road, Cambridge CB2 2QH.

1. Bender, W., Akam, M., Karch, F., Beachy, P., Pfeifer, M., Spierer, P., Lewis, E.B. and Hogness, D.S. *Science* **221**, 23 (1983).
2. Hafen, E., Levine, M., Garber, R.L., & Gehring, W.J. *EMBO J.* **2**, 617 (1983).
3. Akam, M. *EMBO J.* (in the press).
4. Greenwald, I.S., Sternberg, P.W. & Horvitz, H.R. *Cell* **34**, 435 (1983).

100 years ago

SOME studies pursued by the writer have led him to believe that in the atom groupings which modern chemistry reveals to us the several atoms occupy distinct portions of space and do not lose their individuality. The object of the present paper is to show how far this conclusion is in harmony with the symmetrical forms of crystals.

If we are to suppose that crystals are built up of minute masses of different elements symmetrically disposed, it is natural to inquire what very symmetrical arrangements of points or particles in space are possible. It would appear that there are but five (Fig. 2-6).



From *Nature* **29**, 186 (1883).

Ecology

The decline of the chough

from Peter D. Moore

THE results of a new survey¹ documenting the decline of one of western Europe's cliff birds, the chough (*Pyrrhocorax pyrrhocorax*), provides a particularly clear example of just how complex and unpredictable are the consequences of changes in land use. The decline of the chough has most probably resulted from a reduction in grazing pressure — including changes from sheep to cattle grazing — on maritime grasslands which have allowed the growth of taller and denser swards of grass.

The survey¹, by Bullock, Drewett and Mickleburgh, shows that the present chough population (about 900 breeding pairs plus about 800 non-breeding birds in the British Isles) is similar to that recorded in the last survey in 1963, but that the geographical range of the species has greatly contracted since the early nineteenth century. Then it bred in the Channel Islands, along the south coast of Britain from Kent to Cornwall and along most of our rocky western coasts. Now it is limited to west Wales, the Isle of Man, Jura, Islay and Colonsay in Scotland, although it is still widespread in western and southern Ireland.

Conservationists will be relieved to know that the decline of the chough has halted, but the causes of the decline must be understood if the bird's future is to be secured. Human persecution has probably played a part in the decline, but other corvids have been subjected to even greater pressures without succumbing in this way. Habitat disturbance is not likely to have been a serious factor, since these birds are not of a nervous disposition and their nest sites on cliffs are usually inaccessible. Climate may be a factor which confines the chough to maritime areas in northern Europe; it is not so confined in southern Europe and in Asia. Bullock and colleagues show that the highest adult mortality is in February, when frost may limit the availability of the soil invertebrates upon which it largely feeds. Where it occurs inland in Asia, such as in the high plateau of northern Iran, it spends its winter mainly in areas where frost is not too severe, and often in the vicinity of villages.

But it is the feeding studies of Bullock and his colleagues which seem to hold the key to the chough's decline. The chough's preferred feeding habitat is short maritime-cliff grassland, followed by short turf on dunes, then by rough (unimproved) grassland. Arable, bog, heathland and improved grassland are neglected by coastal choughs. The preference for short grass can be explained by two factors: the greater ease with which approaching predators can be seen² and its suitability for the chough's feeding habits. Arable and bog habitats

either do not contain the right kind of food, or the chough finds it hard to obtain because it has a long bill and feeds by probing relatively deeply into the soil. Preferred food items are tipulid larvae, ants, coleopteran larvae and spiders, suggesting that it is feeding technique rather than the simple availability of food that is important.

Improvement of grassland by fertilization produces a taller sward of higher productivity but lower species density³ and will have an adverse effect on the chough population. Sheep-grazed

grassland will be preferable to that grazed by cattle, as sheep provide a closer-cropped turf. Just how directly sheep and chough densities are related can be seen on a small island, the Calf of Man. When farming was stopped the chough breeding population fell by 60 per cent over the next ten years. Only when sheep were reintroduced did the chough population gradually return to its former level. □

Peter D. Moore is Senior Lecturer in the Department of Plant Sciences, King's College, 68 Half Moon Lane, London SE24 9JF.

1. Bullock, I.D., Drewett, D.R. & Mickleburgh, S.P. *Br. Birds* 76, 377 (1983).
2. Brough, T. & Bridgman, C.J. *J. appl. Ecol.* 17, 243 (1980).
3. Grime, J.P. *Plant Strategies and Vegetation Processes* (Wiley, Chichester, 1979).

Palaeontology

Archaeopteryx — no new finds after all

from Michael Howgate

THE rumours of a new find of the earliest known fossil bird, *Archaeopteryx*, reported in *News*¹ last year, have proved to be unfounded. When, at last, the two new specimens were examined by German palaeontologists, one was found to be the remains of a pterosaur (a not unforgivable misidentification since the first *Archaeopteryx* ever discovered was referred to a new species of *Pterodactylus* in 1855 and the mistake was not recognized until 1970²); the other, however, somewhat more strangely, turned out to be a fish. So the plethora of current theories³ will have to be debated over the bones of the five known skeletons.

The four day symposium devoted entirely to *Archaeopteryx* being organised by the Jura museum in Eichstatt (in September of next year) seems an appropriate venue for a thorough re-examination of this important fossil. Unfortunately, while the symposium is already oversubscribed with protagonists of the various theories, the subject of their deliberations, the actual specimens, will be in short supply.

Only the Eichstatt specimen, which is a resident feature of the Jura museum, is certain to be present. Of the other specimens, the Munich feather imprint and the very incomplete Tyler museum specimen will possibly be in attendance. The Maxberg (also known as Solenhofen and sometimes Erlangen) specimen was scheduled to attend but the owner, a local private collector, has since changed his mind — negotiations are continuing however.

The key Berlin specimen has been requested and the application has gone to a



The supposed break (arrowed) in the third topographic digit of the right manus of the Berlin specimen.

higher authority than the Humboldt University Museum for consideration, but the outlook is not very hopeful. As far as the London specimen is concerned, a decision has already been made at the British Museum (Natural History): their specimen will be staying at home. However a detailed video film may be made of the specimen to be shown at the symposium.

It seems a great pity that, with all the expense involved and the time of so many scientists, it is impossible to have all the specimens on hand to adjudicate, on the spot, between the theories. The symposium would provide an ideal opportunity to iron out, at the very least, some of the contradictory interpretations of the osteology on

which the contending theories rest — for example, the orientation of the pubic bone and the supposed breakage in the topographic third digit of the manus.

The latter problem has been discussed in several recent papers by proponents and opponents of a theropod ancestry for *Archaeopteryx*. Tarsitano and Hecht^{4,5} resurrected the broken phalanx hypothesis of Heinroth⁶ and arrived at a derived, non-theropod phalangeal formula (2.3.3.) as against the generally accepted, more primitive and theropod-looking (2.3.4.) formula. In order to explain the break these authors originally suggested that it was caused during fossilization by pressure from a flange of bone on the second topographic digit, as seen on the left manus of the Berlin specimen. This flange supposedly served to brace and strengthen the wing. On the basis of my own examination of the specimen, the 'flange' appears to be a secondary outgrowth, due to injury or disease, and not an original morphological structure. In any event, as Thulborn and Hamley⁷ have noted, there is no indication of a similar outgrowth or flange opposite the 'breakage' on the right manus of the Berlin specimen or any of the other specimens.

The breakage and twisting of the third topographic digit (all the specimens of *Archaeopteryx* with a well preserved manus have the second and third fingers crossed) is variously assumed by Tarsitano and Hecht⁴ to be the result of 'death position, preservation or mode of death'; and, more recently, as 'preservation or stalling in flight, falling and impact of wing as it hit the sea surface'. This conjures up a scenario of *Archaeopteryx*, sweeping out over the Solenhofen lagoon, stalling and crossing its fingers hopefully as it plunges at high speed into the sea wing-tips first, breaking exactly the same digit of each hand in exactly the same place every time. In fact, as a detailed photograph shows (see plate), the supposed break is a good joint, with the more distal phalanx rotated through 90° to reveal the tendon groove as described by Thulborn and Hamley⁷.

Ideally, of course, many of these problems could be cleared up if one of the specimens was prepared out fully in acid — a procedure which a guardian of one of the specimens recommended be tried out, but on someone else's specimen. A start could, however, be made if theorists and specimens could be closeted together in Germany for four days next year. □

Michael Howgate is in the Department of Zoology, University College, London, Gower Street, London WC1E 6BT.

Genetics

Duplex breaks in DNA as recombination initiators

from Harold L.K. Whitehouse

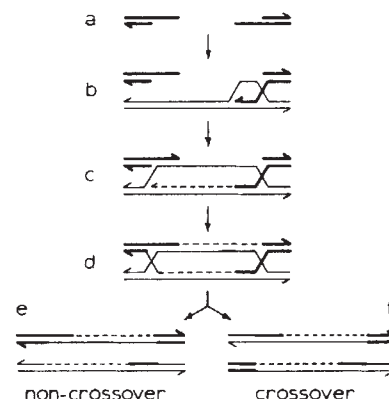
It is well known that double-strand breaks in DNA are often lethal: the natural function of restriction enzymes is to make breaks in foreign DNA to prevent its entry. Duplex breaks can, however, be repaired in the DNA of the yeast *Saccharomyces cerevisiae* if the *RAD52* gene is present¹. This discovery points to an enzymic mechanism for repair of breaks using an intact sister molecule. A mechanism for genetic recombination has now been proposed that uses duplex breaks as initiation sites²; it is of particular importance because it can also provide a new model for gene conversion.

The advent of techniques for yeast transformation led to the discovery that a duplex gap is repaired in the course of integration of the DNA into a homologous molecule and that the *RAD52* gene is needed for integration³. The molecule containing a gap was a plasmid derived from *Escherichia coli* that contained part of a yeast chromosome from which several hundred base pairs had been removed by restriction enzymes. The two ends of the linear molecule were thus homologous with regions a short distance apart in the appropriate yeast chromosome. Orr-Weaver *et al.*³ suggested that the 3' ends of the plasmid may invade the homologous regions of the chromosome, the ends then priming repair synthesis in which the *RAD52* product perhaps plays a part. Nicking of single-stranded regions that would arise when the repair synthesis extended beyond the original gap would produce the observed result: integration accompanied by duplex gap repair.

Repair of a duplex gap using an intact sister molecule may have wide significance because it provides an explanation for gene conversion. The favoured hypothesis to account for gene conversion is the correction of mismatched bases in heteroduplex DNA by the excision of one strand followed by the resynthesis of a strand complementary to the complete remaining one. Szostak *et al.*² now propose a quite different double-strand-break repair model for recombination, based on the yeast transformation studies. It was already known that the *RAD52* gene product is required for meiotic and mitotic gene conversion^{4,5}.

In support of their hypothesis, Orr-Weaver and Szostak⁶ have found that duplex-break repair, like conversion, is associated with crossing over in about 50 per cent of cases. The evidence was produced as follows. Sixty base pairs within the yeast *HIS3* gene, which was carried by the plasmid, were removed with

a restriction enzyme. The corresponding gene in the yeast chromosome carried a mutation that mapped outside the region excised in the plasmid, so that an intact *HIS3* gene could be produced in the plasmid only following duplex gap repair. Thus, repair was signalled by an ability of the yeast cells to grow in the absence of histidine. Integration of the plasmid into the yeast chromosome was indicated by the stability of the histidine prototrophs resulting from the repair process. Integration implies that reciprocal exchange has taken place. In its absence, histidine independence will be lost from some of the daughter cells. Approximately



The five stages of the double-strand-break repair model described by Szostak *et al.* A double-strand cut is made in one duplex and a gap flanked by 3' single strands formed by exonucleases, and then repaired by the process described below. From Szostak *et al.* (ref.2, p.30).

equal numbers of stable and unstable histidine prototrophs were obtained, implying 50 per cent crossing over associated with the duplex repair. Similar results were obtained with the repair of other duplex gaps in *HIS3* produced with other restriction enzymes.

It was still possible that the repaired but not integrated plasmids arose secondarily, all gap repair proceeding through integration. This was tested using an integrated plasmid containing the yeast *URA3* gene which conferred uracil independence⁶. Excision of the plasmid would cause instability in the character, but all transformants for a second plasmid that were tested were found to be stable for uracil prototrophy.

Additional experiments were carried out⁶ in which correct repair of a duplex gap was not forced, as it is if histidine prototrophs are selected. The technique was to

1. Howgate, M. *Nature* 300, 469 (1982).
2. Ostrom, J.H. *Science* 170, 537 (1970).
3. Benton, M. *Nature* 305, 99 (1983).
4. Tarsitano, S. & Hecht, M.K. *J. Linn. Soc. Zool.* 69, 149 (1979).
5. Hecht, M.K. & Tarsitano, S. *Geobios mem. spec.* 6, 141 (1982).
6. Heinroth, O. *J. für Ornithologie* 71, 277 (1923).
7. Thulborn, R.A. & Hamley, T.L. *Aust. J. Zool.* 30, 611 (1982).

use a plasmid carrying *URA3* and to select for uracil-independent transformants. It was then found that partial repair of the gap in *HIS3* sometimes occurred, probably resulting from enlargement of the gap beyond the region of homology to yeast DNA. Gene conversion was also observed in DNA flanking a gap. Thus, transformants were occasionally found that had a functional chromosomal *HIS3* gene but an unstable, that is, non-integrated, plasmid. This derivation of *HIS3*⁺ without gap repair implied transfer from plasmid to chromosome of the normal allele of the chromosomal *his3* mutation, and was attributed to the formation of heteroduplex DNA and subsequent mismatch repair. On the other hand, several experiments indicated that linear plasmid DNA could be degraded and then the enlarged gap repaired. It was evident that conversion in regions flanking the original gap might occur by either mechanism.

The recombination model proposed by Szostak *et al.*² postulates that recombination is initiated by a duplex break, enlarged to form a gap. The gap is repaired by the process already outlined. This results in conversion for mutant sites within the gap and in heteroduplex DNA for mutant sites flanking the gap. Szostak *et al.*² propose that events in these heteroduplex regions follow the Meselson-Radding model⁷. So the new hypothesis can be visualised as a region of duplex

repair flanked on each side by regions conforming to the predictions of the Meselson-Radding model.

As Szostak *et al.*² point out, their hypothesis can explain why yeast mutants, including large deletions and insertions, show parity in frequency of conversion to wild type and to mutant: they suggest that the duplex gap is much enlarged, so that most conversion in yeast results from gap repair rather than mismatch correction. Parity would arise if there was an equal chance of a chromatid of either parentage suffering a duplex break. The hypothesis also explains why recombination initiation sites function as a recipient of genetic information from the other recombining chromatid rather than as a donor of information to it. There is evidence for this from three sources: the M26 mutant in *Schizosaccharomyces pombe*⁸, *cog* in *Neurospora crassa*⁹ and YS17 in *Sordaria brevicollis*¹⁰ (for review see ref. 11).

For many years Stahl¹² has believed that conversion in yeast arises by local DNA synthesis not triggered by mismatched bases. The duplex-break repair model², by combining this hypothesis with that of mismatch correction, surmounts the objections¹³ to Stahl's hypothesis but at the price of a model that will be difficult to test. There are two reasons for this. In the first place, many of its predictions are the same as those of the Meselson-Radding model⁷, some aspects of which already have substantial support^{11,14}. Second, the

patterns of polarity in recombination within eukaryotic genes have led to the conclusion that recombination is normally initiated outside the genes. The postulated duplex breaks are therefore expected to occur in regions lacking genetic markers. Szostak *et al.*² suggest that a direct search be made for double-strand breaks arising during the period of commitment to meiotic recombination. □

Harold L.K. Whitehouse is at the Botany School, University of Cambridge, Downing Street, Cambridge CB2 3GA.

1. Resnick, M. & Martin, P. *Molec. gen. Genet.* **143**, 119 (1976).
2. Szostak, J.W., Orr-Weaver, T.L., Rothstein, R.J. & Stahl, F.W. *Cell* **33**, 25 (1983).
3. Orr-Weaver, T.L., Szostak, J.W. & Rothstein, R.J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6354 (1981).
4. Prakash, S., Prakash, L., Burke, W. & Montelone, B. *Genetics* **91**, 31 (1980).
5. Game, J., Zamb, T., Braun, R., Resnick, M. & Roth, R. *Genetics* **91**, 51 (1980).
6. Orr-Weaver, T.L. & Szostak, J.W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4417 (1983).
7. Meselson, M. & Radding, C. *Proc. natn. Acad. Sci. U.S.A.* **72**, 358 (1975).
8. Gutz, H. *Genetics* **69**, 317 (1971).
9. Catcheside, D.G. *The Genetics of Recombination* (Arnold, London, 1977).
10. MacDonald, M.V. & Whitehouse, H.L.K. *Genet. Res.* **34**, 87 (1979); *Genet. Res.* **41**, 155 (1983).
11. Whitehouse, H.L.K. *Genetic Recombination: understanding the mechanisms* (Wiley, Chichester & New York, 1982).
12. Stahl, F.W. *Genetics* **61**, Suppl. 1 (1969); *Genetic Recombination: thinking about it in Phage and Fungi* (Freeman, San Francisco, 1979).
13. Holliday, R. & Whitehouse, H.L.K. *Molec. gen. Genet.* **107**, 85 (1970).
14. Paquette, N. & Rossignol, J.-L. *Molec. gen. Genet.* **163**, 313 (1978).

Astronomy

IRAS circulars 4 & 5

The source name consists of four parts: (1) the letters 'IRAS' to indicate the origin; (2) the right ascension (RA) in hours and minutes, seconds omitted; (3) declination (Dec) in decimal degrees, multiplied by 10 and then truncated (thus +32° 42.3 becomes +327); (4) an appendix starting with 'P' and followed by the number of the circular; this appendix stresses that the data are preliminary. Position is given at Equinox 1950.0.

Source IRAS	RA (1950) h min s	Dec (1950) deg arc min	Flux density (Jy)				Source IRAS	RA (1950) h min s	Dec (1950) deg arc min	Flux density (Jy)			
			12 µm	24 µm	60 µm	100 µm				12 µm	25 µm	60 µm	100 µm
1608-185P04	16 08 38	-18 30.7	8.7	15	22	22	1650+024P04	16 50 28	+02 29.0	0.51	3.7	26	34
1623+030P04	16 23 33	+03 01.2	<0.2	0.66	3.8	5.5	1651+305P04	16 51 41	+30 31.0	<0.2	<0.3	1.9	4.1
1624+116P04	16 24 25	+11 41.5	<0.3	<0.4	2.6	7.2	1700-234P04	17 00 40	-23 28.6	4.8	6.6	1.9	<3
1626+037P04	16 26 13	+03 43.4	<0.2	<0.3	2.2	3.6	1705-022P04	17 05 33	-02 16.5	6.6	6.1	1.2	<2
1627+031P04	16 27 49	+03 07.4	<0.2	<0.3	1.7	2.5	1709-165P04	17 09 22	-16 33.5	8.7	7.8	0.99	<3
1628+041P04	16 28 27	+04 11.4	<0.3	0.99	7.8	16	1710-032P04	17 10 14	-03 12.5	<0.3	1.5	3.0	<2
1639-096P04	16 39 56	-09 37.6	<0.4	1.1	8.7	16	1713-102P04	17 13 50	-10 17.5	0.57	2.2	19	31
1640-188P04	16 40 58	-18 51.7	<0.2	4.4	4.2	<3	1717-087P04	17 17 09	-08 44.0	38	34	6.0	<3
1645+033P04	16 45 28	+03 23.5	<0.2	<0.3	2.2	3.5	1718+113P04	17 18 02	+11 22.0	<0.2	0.40	2.3	3.7
1647-113P04	16 47 37	-11 22.9	1.6	5.2	2.7	<4	1720+129P04	17 20 49	+12 57.1	<0.4	<0.2	1.9	3.2

Source IRAS	RA (1950) h min s	Dec (1950) deg arc min	Flux density (Jy)				Source IRAS	RA (1950) h min s	Dec (1950) deg arc min	Flux density (Jy)			
			12 µm	25 µm	60 µm	100 µm				12 µm	25 µm	60 µm	100 µm
0441+727P05	04 41 52	+72 46.2	<0.3	1.2	4.6	6.8	0531-219P05	05 31 13	-21 58.8	0.42	0.77	9.7	32
0449+781P05	04 49 44	+78 06.6	<0.6	0.64	6.6	12	0533+541P05	05 33 45	+54 08.0	<0.2	0.49	5.4	8.3
0506+536P05	05 06 07	+53 38.7	0.34	1.7	9.4	16	0536+467P05	05 36 09	+46 44.2	170	200	77	34
0507+471P05	05 07 00	+47 07.0	0.58	3.0	17	38	0538-220P05	05 38 06	-22 01.7	<0.2	<0.2	2.1	4.2
0507+528P05	05 07 19	+52 48.9	200	290	69	32	0540-240P05	05 40 57	-24 05.2	<0.3	0.52	2.8	4.4
0508+796P05	05 08 16	+79 36.7	<0.3	0.62	6.0	11	0541+586P05	05 41 24	+58 40.8	0.60	0.87	16	40
0512+531P05	05 12 52	+53 08.2	<0.4	0.67	3.3	6.6	0547-303P05	05 47 47	-30 18.7	<0.2	<0.3	3.7	8.3
0512+514P05	05 12 59	+51 28.7	<0.3	1.0	7.2	9.0	0552-327P05	05 52 01	-32 45.1	<0.2	<0.4	1.8	4.2
0513+581P05	05 13 28	+58 11.1	<0.3	0.47	5.2	13	0600+477P05	06 00 22	+47 47.9	34	31	5.1	<10
0516+432P05	05 16 39	+43 15.3	0.33	0.79	6.3	11	0610+668P05	06 10 39	+66 51.2	<0.3	<0.4	3.8	8.6
0517+428P05	05 17 17	+42 49.8	0.58	0.73	4.5	14	0623+744P05	06 23 57	+74 28.6	<0.2	0.88	5.4	8.3
0522+416P05	05 22 07	+41 39.2	2.6	18	140	190	0705+719P05	07 05 32	+71 55.0	<0.2	<0.3	2.4	6.1
							0706+718P05	07 06 45	+71 50.0	<0.4	0.42	4.1	10

High-resolution observations of ionized gas in central 3 parsecs of the Galaxy: possible evidence for infall

K. Y. Lo & M. J. Claussen

Owens Valley Radio Observatory, California Institute of Technology, HM Robinson Laboratory of Astrophysics, Pasadena, California 91125, USA

A 1 arc s resolution radio map shows a wealth of detail in the 'spiral' distribution of ionized gas within 1.5 pc of the galactic centre. Three streams of ionized gas passing near the centre can be identified. These streams could be due to molecular gas falling in towards the centre, in which case we have the first direct observation of the paths of accretion by a galactic nucleus.

THE centre of our Galaxy has long been an object of interest to astronomers. At a distance of 10 kpc, it allows observations of a galactic nucleus at an unsurpassed linear resolution. Observations of the galactic centre could help in understanding the active phenomena found in the nuclei of external galaxies.

In terms of energetics, the activities in our galactic centre are mild compared with quasars, Seyfert galaxies and radio galaxies. However, the centre is interesting because of a variety of phenomena which suggest past or present activities: the large-scale expanding motions of neutral gas observed towards the galactic centre¹ may indicate an energetic explosive event having taken place about 10^7 yr ago. The presence of a compact non-thermal radio source ($d \leq 100$ AU) with properties similar to those of extragalactic compact radio sources² is a possible signature of a massive black hole³. A central point mass is also suggested by the large velocities of the ionized gas within the central cubic parsec⁴.

Recently, observations have been reported of a time-variable e^+e^- annihilation γ -ray line from the galactic centre, indicating a compact source of positrons⁵, and of a $2.06 \mu\text{m}$ He(I) line with a linewidth of $1,500 \text{ km s}^{-1}$ originating from IRS16 (ref. 6), the presumed core of the central star cluster⁷.

Observations of the ionized gas at the galactic centre are important because they provide information on physical conditions such as the velocity field and the luminosity of the ionizing sources. However, such observations are restricted to IR and radio wavelengths because of high obscuration along the line of sight. The first reliable radio aperture synthesis map⁸ (6×30 arc s resolution) of Sgr A, the radio source at the galactic centre, clearly resolved it into two components: Sgr A East, presumably a supernova remnant, and Sgr A West, a region of ionized gas with the compact radio source² at the centre.

Lacy *et al.*^{4,9,10} mapped the $12.8\text{-}\mu\text{m}$ Ne(II) emission line from Sgr A West with ~ 3 arc s resolution. They identified 14 ionized gas clouds within ~ 1.5 pc of the centre, with large radial velocities (up to $\pm 260 \text{ km s}^{-1}$) and large internal velocity dispersions ($\geq 50 \text{ km s}^{-1}$). The inferred lifetime of the clouds is very short ($10^3\text{--}10^4$ y), which poses problems for explaining the origin and fate of the ionized gas^{9,10}. In addition, the velocity field is complicated and no simple pattern is apparent⁴. Radio recombination line measurements¹¹ show that the Ne(II) emission regions can be identified with the thermal radio emission.

With the advent of the Very Large Array (VLA) telescope of the National Radio Astronomy Observatory (NRAO) high-resolution radio observations of the ionized gas in the galactic centre have become possible. Brown, Johnston and Lo¹² used the partially completed VLA to obtain a 2×8 arc s resolution map of Sgr A West, which showed for the first time the unusual

'ridge-like' distribution of the ionized gas. More recently, observations of Sgr A at several wavelengths by Ekers *et al.*¹³ revealed that the ionized gas at the galactic centre is distributed in an intriguing 'spiral' pattern. They considered various explanations and concluded: "We have not been able to definitely exclude any of the models discussed, but the tidal distortion of infalling material seems to provide the most coherent explanation of the observed phenomena".

We present below a 6-cm VLA map of the ionized gas within the central 3 pc at 1 arc s resolution. It adds considerable detail to the previous maps. In addition, we offer a reinterpretation of the Ne(II) line profiles to suggest an explanation of the velocity field of the ionized gas. Finally, we present a model to explain the origin and appearance of the ionized gas, and other observations of the galactic centre.

Observations and data reduction

The 6-cm observations reported here were made with the VLA. Sgr A West was observed in January 1982 with the array in the 'C' configuration (but with a long north arm), and again in October 1982 with the array in the standard 'B' configuration¹⁴.

Both data sets were first calibrated relative to the radio source 1730-130 (NRAO 530). The flux density scale was determined relative to 3C286 (with an adopted flux density of 7.4 Jy at 6 cm). With the compact nonthermal source as an internal reference, the self-calibration procedure¹⁵ was applied to each data set separately. The visibility data from the two configurations were then combined, and maps were made with a resolution of 1×1.3 arc s ($\alpha \times \delta$). The CLEAN algorithm¹⁶ was used to correct for imperfect coverage of the interferometer spacings.

Because of the high resolution, the map is not very sensitive to extended low surface brightness features. Here we shall be concerned primarily with the high brightness details.

Results

Figure 1 shows a contour map of Sgr A West and Fig. 2 a radiograph with the compact radio source subtracted. The contour map provides a quantitative display of the brightness levels, while the radiograph gives a better impression of the ionized gas distribution.

Figures 1 and 2 show that the distribution of ionized gas within 1.5 pc of the galactic centre is complicated. The overall distribution is dominated by three bright arms (for convenience, we use the term 'arms' to refer to elongated features in the ionized gas) that pass near the centre. As the dynamical centre of the galaxy is not defined at the 1 arc s level, we shall use here the term 'centre' to refer to the region immediately surrounding, and including, the compact radio source.

Features of lower surface brightness ($\sim$ four times lower than the bright arms) are found on the periphery of the distribution: the south arm, the northern continuation of the east arm, and an incomplete loop starting about halfway up the north arm, going north-east, bending around past the northern continuation of the east arm, and ending near the centre (see Fig. 3). Additional features exist, perhaps more evident in the radiograph.

The ionized gas is very clumpy and patchy. Compact high brightness clumps are found along the three bright arms (see Fig. 1); almost all the compact structures there can be identified with the features in the 10- μ m continuum map of Becklin *et al.*⁷. These small-scale structures are embedded in the diffuse background of the arms, so that it is difficult to delineate their extent. However, the size scales of the fine structure appear to be ≤ 3 arc s (0.15 pc) and could be as small as ~ 1 arc s in a few cases. 'Spurs' of ionized gas emanate from these high-brightness clumps, normal to the arms. These are especially obvious on the south side of the east and west arms.

The south arm has a lower surface brightness than the three bright arms. As evident in Fig. 2, it consists of at least four major substructures. Where the south and west arms apparently meet, there is a break as well as a large difference in the brightness of the two arms. (In addition, there is a velocity discontinuity between the two arms at that point, see below.) While not obvious in lower resolution maps, the new map indicates that the south arm is not connected to the west arm but continues northward toward the tip of the north arm. The previous interpretation of the north, west and south arms as parts of a 'spiral' distribution¹³ is therefore no longer compelling.

The electron density of the ionized gas can be derived from the brightness temperature. Following the procedure of Brown *et al.*¹², we obtain $n_e \approx 5 \times 10^4 \text{ cm}^{-3}$ for the high brightness clumps, and $n_e \approx 10^3 \text{ cm}^{-3}$ for the lower surface brightness features. The total mass of ionized gas in the arms is $\approx 60 M_\odot$ and is distributed as follows: $\sim 10 M_\odot$ each in the north and east arms, $\sim 5 M_\odot$ in the west arm, and $\sim 35 M_\odot$ in the south arm.

As evident in Fig. 1, the compact radio source is not situated at any peak of the ionized gas (see for example refs 13, 17) and is offset by ~ 2 arc s (0.1 pc) to the north of the ionized gas. However, the compact source still appears to be located roughly at the centroid of the ionized gas distribution.

Velocity field of the ionized gas

The velocity field of the ionized gas is given by the 12.8- μ m Ne(II) emission line observations of Lacy *et al.*¹⁰. Figure 3 shows the radial velocity of the gas at various positions. The systematic variation in the velocity along the arms, especially at points away from the centre, is obvious. Within the central 10 arc s, the line profiles are complicated. Lacy *et al.*¹⁰ described the motion of the gas as a combination of rotation and random motion. While rotation may explain the general red- and blue-shift in the emission line east and west of the centre, respectively, the presence of both large blue- and redshift in the line profiles near the centre is inconsistent with rotation.

The high-resolution map of the ionized gas suggests that the three bright arms originate from near the compact radio source, where the Ne(II) spectra are complicated. This led us to re-inspect the Ne(II) profiles within the central 30 arc s to determine if they could be decomposed into three velocity components. The decomposition was done by inspection, assigning a velocity component to each of the local maxima in the spectra. Figure 4 shows the spectra within the central 10 arc s and the decomposed velocity components. The velocity of each component appears to vary continuously with position.

In Fig. 5, we have superposed the decomposed line profiles onto a contour map of the central region of the ionized gas. If we follow the same velocity component from point to point in the map, we can identify a spatially continuous stream of gas. There are three streams which correspond to the three bright arms. Thus it appears that the systematic variation of velocity along the outer parts of the arms continue into the centre, and

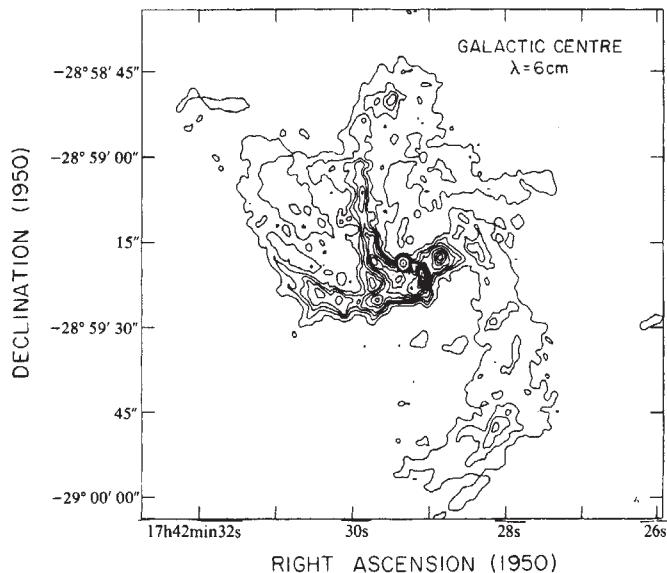


Fig. 1 VLA map of Sgr A West. The compact nonthermal radio source is at the map centre at the nominal position of RA(1950) = 17 h 42 min 29.335 s, dec.(1950) = $-28^\circ 25' 18.6''$. Because of self-calibration, the relative position of any feature on the map to the compact source is highly accurate, whereas the absolute position is probably only ± 0.1 arc s. The contours are at 0.8, 1.6, 2.3, 3, 4, 5, 6, 7, 8, 9, 10, 50 and 95 times 10.1 mJy per beam or 329 K. The half-power beamwidth of the restoring beam is 1.1×1.3 arc s, as indicated by the second to the last contour of the compact source. 15 arc s corresponds to 0.75 pc at a distance of 10 kpc.

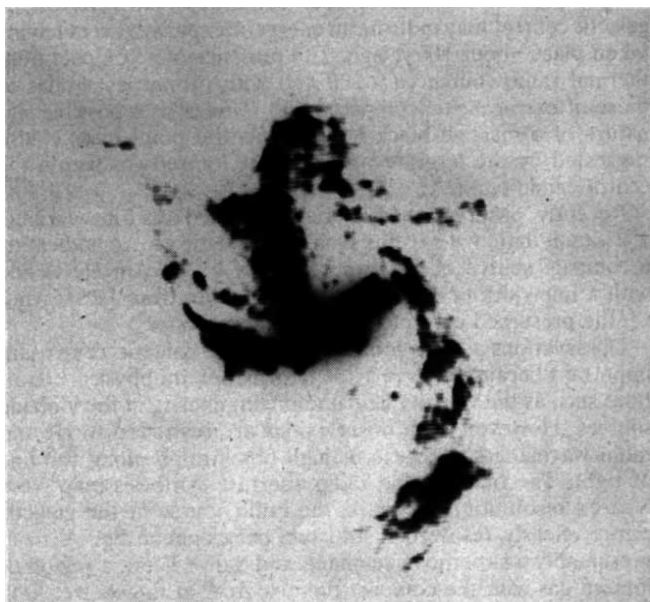


Fig. 2 Radiograph of Sgr A West. Same as Fig. 1 except that the compact nonthermal radio source at the map centre has been subtracted.

we can trace three streams of gas that appear to converge to or diverge from near the centre. Close inspection of Fig. 5 shows, however, that the three streams of ionized gas do not meet at a single point.

In Fig. 3, we have drawn lines to delineate the loci of contiguous velocities in the ionized gas. The lines also outline the current position of the ionized gas clouds along their trajectories and suggest the paths of the gas streams. The overall organization of the ionized gas may also be indicated by the existence of certain symmetries in the velocity field: the south and the north arms have velocities of opposite signs, which is also true for the east and the west arms. However, for the north and the

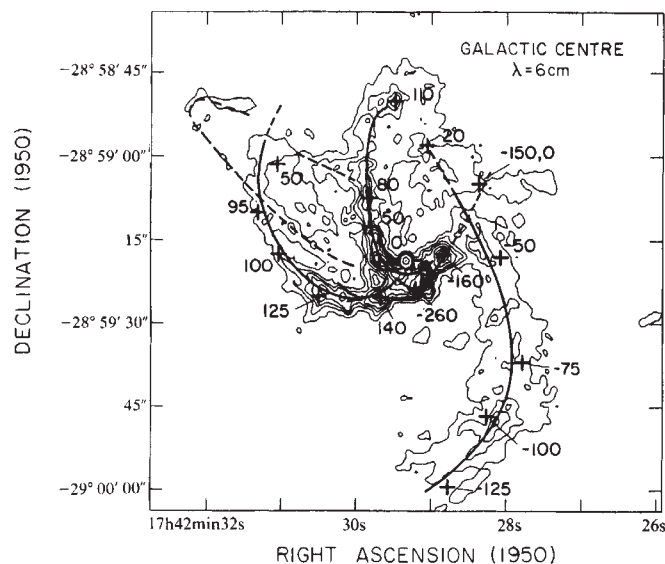


Fig. 3 This shows the overall velocity field of the ionized gas in Sgr A West. The numbers next to the crosses are the radial velocity (in km s^{-1}) of the Ne(II) lines measured at those positions⁴. Lines are drawn along the loci of contiguous velocities to suggest the current position of the ionized gas along their trajectories. Multi-wavelength observations of Ekers *et al.*¹² indicate that the south arm may lie in front of the whole complex of ionized gas. Observations of H(I) absorption¹⁸ and of $[\text{OI}]$ (ref. 19) suggest that the surrounding molecular gas is rotating about the centre in the sense that it is approaching us south of Sgr A West, moving northward in front of Sgr A West and receding from us to the north. The south and the north arms could be the ionized inner edge of the orbiting molecular gas.

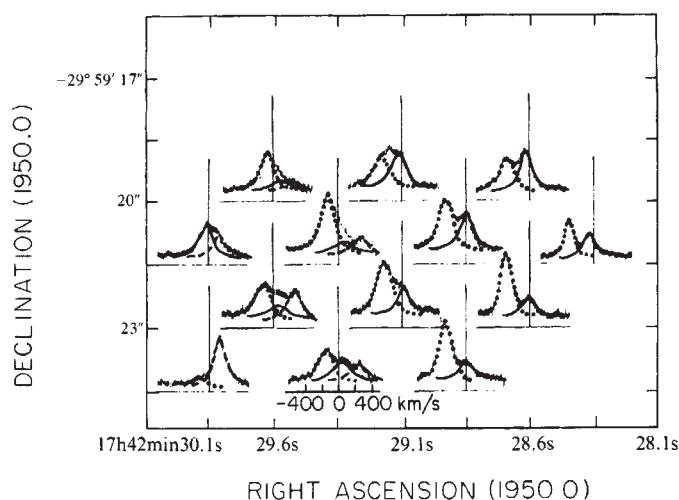


Fig. 4 This shows the more complicated $12.8 \mu\text{m}$ Ne(II) line profiles at locations near the compact radio source. Superposed on them are the components obtained by decomposing the Ne(II) profiles by inspection, which suggests that these Ne(II) lines consist of three components at different velocities. The three components are denoted by dotted, dashed and solid lines.

south arms, the magnitude of the velocity decreases towards the centre, whereas for the east and the west arms, the largest velocities are seen near the centre.

As mentioned above, the new map suggests that the south arm is not connected to the west arm. This is further suggested by the large velocity discontinuity where the two arms appear to meet (see Fig. 3). If the gas in the south arm is moving from

south to north, the negative velocities would imply that the south arm is in the foreground of the centre. This is consistent with the absorption of the diffuse emission by the south arm, as noted by Ekers *et al.*¹³. It appears convincing, then, that the south arm continues northward past the west arm, and is connected to the north arm, forming a single feature.

The symmetry in the velocity field thus suggests that the north and the south arms form a single feature that is orbiting about the centre. The east and the west arms, on the other hand, could not be part of the same structure because of the large velocity difference and must be independent features falling towards or being ejected from the centre.

Proposed model of ionized gas

Since the dynamic time scale of the ionized gas is a few times 10^4 yr, the organized structure must be transient. As Ekers *et al.*¹³ pointed out, the explanation for the organized structure of the ionized gas must involve some combination of rotation, ejection or infall of matter. Pure rotation clearly does not work.

Collimated gaseous outflow along twin opposing beams that emanate from a precessing central object as proposed by Brown¹⁷ cannot explain the complicated morphology. The point-reflection symmetry on which the model of precessing twin beams is based is no longer evident in the high resolution map.

If the compactness of the ionized gas clumps is taken as an indication of the expansion time scale, the presence of three streams of gas of comparable age is indicated. Thus, on the ejection hypothesis, three precessing nozzles operating simultaneously in three different directions are required. R. Blandford (personal communication) suggested such a model in which the three streams could be ionized gas ejected along the field lines of a spinning magnetized object. The compact nonthermal radio source is displaced >2 arc s (0.1 pc) to the north of the ionized gas peaks, and Fig. 5 shows that the ends of the three streams of gas are separated by up to 0.5 pc. Generally, models of ejection by a single object cannot explain why the three streams of gas do not meet at a single point.

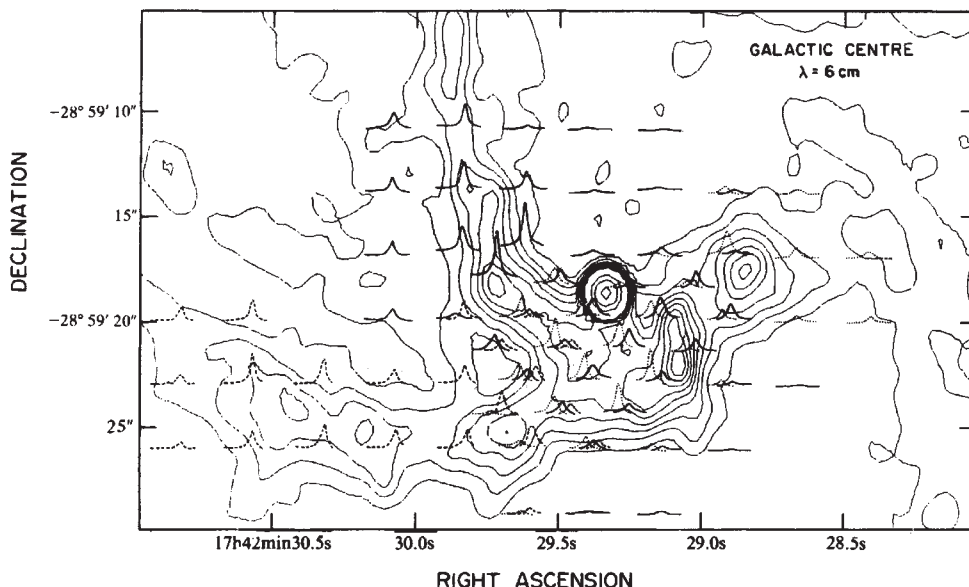
Ekers *et al.*¹³ considered the explanation of the tidal distortion of infalling molecular clouds but could not reconcile it with their interpretation of the structure and the velocity field of the ionized gas. With the identification of the three streams of gas which appear to converge near the centre, the infall model has become more attractive. The Roche density within the central cubic parsec is 10^8 cm^{-3} , much higher than the density of the ionized gas clouds, indicating that tidal effects are very strong.

Far IR continuum observations¹⁸ have shown that there is a deficiency of dust, and presumably neutral gas, in Sgr A West which is surrounded by a 'ring' of material. Liszt *et al.*¹⁹ demonstrated that neutral molecular gas exists just to the north and south of Sgr A West, with the same velocities as the adjacent ionized gas. The south and the north arms could be the ionized inner edge of the surrounding molecular gas which is orbiting about the centre. The higher brightness of the north arm, which may indicate it is closer to the central ionizing source (see below), suggests that at least part of the gas is spiralling inward. The east and west arms are tidal filaments of low angular momentum molecular clouds falling in towards the centre of the Galaxy, along paths roughly normal to the orbital plane of the south and north arms. The dissipation needed to reduce the angular momentum of the infalling gas could be provided by ram pressure of the medium through which the gas is travelling.

The infalling molecular gas probably originates in the immediate vicinity of Sgr A West¹⁵. As the gas clouds come within ~ 1.5 pc of the centre, which is largely clear of gas and dust¹⁸, they are ionized by a central source which is likely to be a cluster of O and B stars¹⁰ formed previously *in situ*. When the gas clouds fall past the O and B stars, even the high density clumps become ionized.

The infall model can explain much of the observations of the ionized gas. The morphology is defined by the trajectories of

Fig. 5 Superimposed on a contour map of the inner portion of the high resolution map of the ionized gas are the three velocity components from the decomposed Ne(II) line profiles (see also Fig. 4). A close inspection of the superimposition suggests that there are three streams of ionized gas contiguous in space and velocity within the central 30 arcs. The dashed-line spectra represent gas at large positive velocities, to be identified with the east arm. The dotted-line spectra represent gas at large negative velocities, identified with the west arm. The solid-line spectra are identified with the north arm.



the infalling gas, so that precise symmetry is not required, and the streams need not meet at a single point. The mass of ionized gas is only about $60 M_{\odot}$, easily available in the molecular clouds in the vicinity of Sgr A West²⁰. Identifying the ionized gas with molecular cloud gas also makes it easy to understand the presence of dust that is well mixed with the gas, as evidenced by the excellent correlation of the $10 \mu\text{m}$ continuum and the 6 cm radio continuum². The large velocity width and complicated profiles of the Ne(II) spectra near the centre are readily explainable by a superposition of the infalling streams of ionized gas. More compact, high density ionized gas clouds are found closer to the centre where the ionization sources are located because the proximity leads to ionization of higher density clumps.

Explanation of related observations

It has been noted by many that the IR properties of the ionized gas clouds at the galactic centre differ from those of the galactic H(II) regions in the following respects: (1) the lack of intrinsic silicate absorption⁷; (2) the peak of the IR energy distribution is at $30 \mu\text{m}$ instead of $100 \mu\text{m}$ (ref. 18); (3) the $10 \mu\text{m}$ to radio continuum ratio is anomalously large^{12,21}; (4) the far-IR luminosity is only slightly greater than the Lyman- α luminosity¹⁸.

All these differences can be ascribed to the absence of large amounts of neutral material (molecular cloud) surrounding the ionized gas¹⁸. This is certainly supported by the far-IR continuum observations of Becklin *et al.*¹⁸ who found a minimum in the dust opacity at Sgr A West. However, it is hard to understand the formation of such high density ionized clouds in a tenuous medium.

Lacy *et al.*^{4,10} considered the ionization of the gas in great detail. The main peculiarity is that $T_{\text{eff}} < 35,000 \text{ K}$ is required by the observations⁴. If the ionization source were stars, there would be a cutoff in spectral types earlier than O8. Such a cutoff in spectral types of the massive stars is possible if the star formation ended $\sim 10^7$ yr ago: main sequence stars earlier than O8 have a lifetime less than 10^7 yr and thus would have evolved by now^{10,21}.

Recent star formation has also been proposed by Lebofsky *et al.*²¹ to explain the high concentration of luminous M supergiants in the centre, the ionization, and the energetics of the galactic centre. The problem, however, has been in understanding why there is recent star formation in the absence of molecular gas in the centre¹⁸, and how the star formation abruptly ended $\sim 10^7$ yr ago.

If we depart from the usual assumption of steady-state conditions, these apparently divergent observations can be understood within the picture that the infall of matter is the latest in

a sequence of events: (1) The O and B stars which now ionize the gas were formed in star formation episodes that ended at least 10^7 yr ago. An additional source of ionization is not excluded. (2) Between $\sim 10^7$ and $\sim 10^4$ yr ago, explosive events must have occurred to disrupt the interstellar medium, so that star formation was halted, the central $\sim 100 \text{ pc}$ became largely deficient in molecular gas²² and Sgr A West was cleared of gas and dust. The latest event must have occurred $\geq 10^4$ yr ago before the start of the infall. (3) Low angular momentum molecular gas is now falling through a cavity towards the centre, is tidally distorted and ionized by the central cluster of OB stars.

Implications

More quantitative analysis of the proposed infall of molecular clouds is clearly needed. Numerical modelling of the morphology and the velocity field of the ionized gas as orbits in a central potential could sort out the relative importance of dissipation, the shape of the potential and the initial angular momentum. This study is being pursued by P. Quinn and G. Susman (personal communication). It may be possible by such simulation to probe the central mass distribution and check the possible existence of a central point mass.

The inferred infall rate, $\sim 10^{-3} M_{\odot} \text{ yr}^{-1}$, is ~ 100 times larger than the required accretion rate onto a $3 \times 10^6 M_{\odot}$ black hole to support the bolometric luminosity of the galactic centre¹⁰. The eventual fate of the infalling gas depends on the conditions near the centre (see ref. 10). The presence of distinct velocity components in the Ne(II) spectra near the centre suggests that the infalling streams of gas have not interacted physically. Presumably, after enough gas has collected in the centre, collisions would randomize the kinetic energy and angular momentum of the infalling gas streams. This would lead to eventual accretion by the central black hole (if one exists). Thus, the infall rate may not equal the accretion rate at any instant.

The transient nature of the ionized gas suggests that the infall of matter towards the centre has not been a steady process in the past. If the luminosity of the galactic centre ultimately depends on the infall rate, variation in the level of activities with time is naturally indicated. This is consistent with the inferred past activities and the proposed picture in the previous section.

The very broad $2.06 \mu\text{m}$ He(I) line⁶ may indicate mass-loss from the centre. For a constant flux of isotropic outflow, the outflow gas density n_e is equal to $3 \times 10^3 \dot{M} (M_{\odot} \text{ yr}^{-1}) / R^2 (\text{pc}) \text{ cm}^{-3}$. Even if fully ionized, the density is too low to detect in the radio continuum unless $\dot{M}$ is $\geq 1 M_{\odot} \text{ yr}^{-1}$.

If the outflow is balanced by infall, up to $10^{-3} M_{\odot} \text{ yr}^{-1}$ can be supported by the inferred infall rate of the ionized gas. Thus, the present radio continuum observations certainly do not preclude isotropic mass loss from the centre that may be taking place.

If our proposed picture is verified, we have for the first time direct observational evidence of infall of material towards the centre of a galaxy. Accretion of interstellar material by a central black hole has been proposed as the energy source for much of the activities seen in quasars, radio galaxies and Seyfert nuclei.

Received 29 August; accepted 1 November 1983.

1. Oort, J. H. *AIP Conf. Proc.* **83**, 180–187 (1982).
2. Lo, K.-Y. *AIP Conf. Proc.* **83**, 1–8 (1982).
3. Rees, M. J. *AIP Conf. Proc.* **83**, 166–176 (1982).
4. Lacy, J. H., Townes, C. H., Geballe, T. R. & Hollenbach, D. J. *Astrophys. J.* **241**, 132–146 (1980).
5. Lingenfelter, R. E. & Ramaty, R. *AIP Conf. Proc.* **83**, 148–159 (1982).
6. Hall, D. N. B., Kleinman, S. G. & Scoville, N. Z. *Astrophys. J. Lett.* **260**, L53–L58 (1982).
7. Becklin, E. E., Matthews, K., Neugebauer, G. & Willner, S. P. *Astrophys. J.* **219**, 121–128 (1978).
8. Ekers, R. D., Goss, W. M., Schwarz, U. J., Downes, D. & Rogstad, D. H. *Astr. Astrophys.* **43**, 159–166 (1975).
9. Lacy, J. H., Baas, F., Townes, C. H. & Geballe, T. R. *Astrophys. J. Lett.* **227**, L17–L20 (1979).

In the centre of our Galaxy, we may be witnessing directly the paths of accretion.

We have benefited from discussions with many people, especially R. Blandford, P. Quinn and P. Goldreich. We thank A. Berkin, R. Brown, K. Johnston and K. Young for help with the observations, and J. Lacy, R. Ekers and J. van Gorkom for providing data before publication. This research was supported by NSF grants AST82-10259 and AST82-10828. The National Radio Astronomy Observatory is operated by the Associated Universities Inc. under contract with the US NSF.

10. Lacy, J. H., Townes, C. H. & Hollenbach, D. J. *Astrophys. J.* **262**, 120–134 (1982).
11. Bergman, J. D. & Schwarz, U. J. *Astr. Astrophys.* **112**, L6–L9 (1982).
12. Brown, R. L., Johnston, K. J. & Lo, K.-Y. *Astrophys. J.* **250**, 155–159 (1981).
13. Ekers, R. D., van Gorkom, J., Schwarz, U. J. & Goss, W. M. *Astr. Astrophys.* **122**, 143–150 (1983).
14. Thompson, A. R., Clark, B. G., Wade, C. M. & Napier, P. J. *Astrophys. J. Suppl.* **44**, 151–167 (1980).
15. Schwab, F. R. *Proc. Soc. Photo-opt. Instrum. Engng.* **231**, 18–24 (1980).
16. Högbom, J. A. *Astr. Astrophys. Suppl. Ser.* **15**, 417–426 (1974).
17. Brown, R. L. *Astrophys. J.* **262**, 110–119 (1982).
18. Becklin, E. E., Gatley, I. & Werner, M. W. *Astrophys. J.* **258**, 135–142 (1982).
19. Liszt, H. B., van der Hulst, Burton, W. B. & Ondreken, M. P. *Astr. Astrophys.* **126**, 341–351 (1983).
20. Genzel, R. *et al.* *Astrophys. J.* (in the press).
21. Lebofsky, M. J., Rieke, G. H. & Tokunaga, A. T. *Astrophys. J.* **263**, 736–740 (1982).
22. Güsten, R. & Downes, D. *Astr. Astrophys.* **87**, 6–19 (1980).

Detection of solar gravity mode oscillations

Philippe Delache* & Philip H. Scherrer

Center for Space Science and Astrophysics, Stanford University, Stanford, California 94305, USA

An analysis of solar velocity data obtained at the Stanford Solar Observatory shows the existence of solar global oscillations in the range 45–105 μHz (160–370 min). These oscillations are interpreted as internal gravity modes of degree $l=1$ and $l=2$. A good estimate of the order of the modes has also been made.

OBSERVATIONS of the global solar velocity field have been recorded¹ at the Stanford Solar Observatory since 1976. Similar recordings from the Crimean Astrophysical Observatory are available² starting from 1974 and a spectral analysis of the combined data showing the prominent 160.01-min peak has been published recently³. A careful examination of the combined Stanford–Crimea spectrum has convinced us that there may be a few other peaks in the low frequency portion of the spectrum which are significantly above the noise level. For that reason, we decided to re-analyse the original Stanford data.

Data set and analysis

The previous report described the observing and data reduction procedures in detail³. The observations, which are differential measures of the line-of-sight velocity of the solar surface, are made by comparing the average Doppler shift from the centre of the solar disk with the Doppler shift from a concentric annulus. As it seemed that interesting features were present in the low frequency part of the spectrum, we extended the computation down to $\nu=0$. Because only the low frequency part of the spectrum is examined here, the data were averaged into 5-min intervals and normalized within each day as in the previous analysis. First the Fourier transform of the 4 yr of Stanford data (1977–80) was computed using a standard fast Fourier transform (FFT) code. It was found that the largest power is in the range 45–105 μHz (about 160–370 min). The resulting power spectrum shows several sets of lines with separations

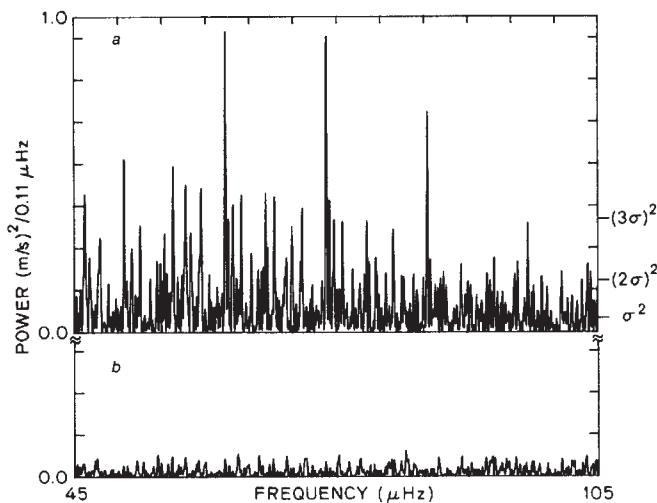


Fig. 1 The power spectrum of velocity observations from the Stanford Solar Observatory in 1979. The spectrum in the range 45–105 μHz (360–160 min) is shown. *a*, The original spectrum; *b*, the spectrum to the same scale after 14 peaks were identified and the associated sinusoidal waves subtracted from the data. The scale shown has been corrected for the average normalizing factor.

corresponding to day side bands (that is, 11.57 μHz) which obviously come from the nightly gaps in the data. Comparing the 4-yr spectrum with individual yearly spectra, we discovered that the strongest lines were more prominent in 1979 than in the other years. This may be due to the relatively clear skies

* Permanent address: Observatoire de Nice, Laboratoire Associé au CNRS 128, BP 252, 06007 Nice Cedex, France.

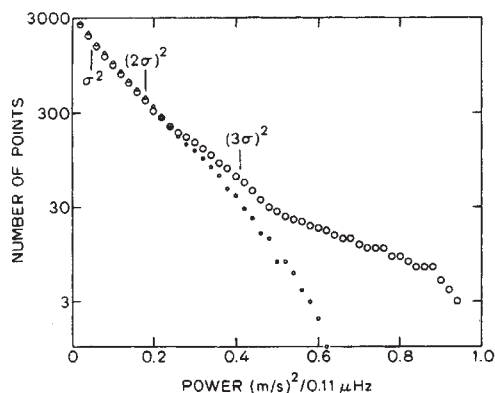


Fig. 2 The cumulative distribution function of the spectrum in Fig. 1a. The log of the number of spectral estimates with size larger than a given value is plotted against that value. The large dots correspond to the observed spectrum. The small dots are from a similar spectrum computed with the data within each day taken in reversed order. The small dots then refer to a spectrum with the same window and noise characteristics but no coherent oscillations. The variance which is deduced from the slope is $\sigma^2 = (20 \text{ cm}^2 \text{ s}^{-2})$.

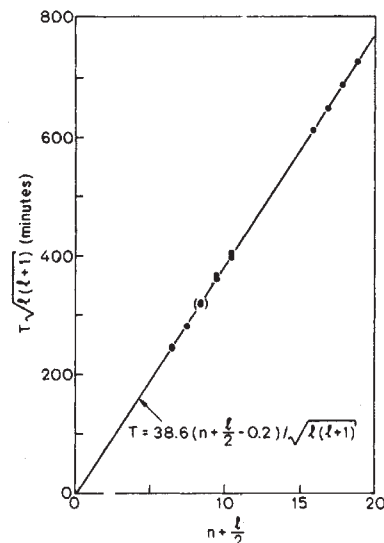


Fig. 3 The observed modes are plotted within the structure of the asymptotic formula described in the text. The line is determined from the three largest peaks that have been assigned.

and the distribution of observing times in 1979. As a first step, we then focus our analysis on that year. Figure 1a shows the 1979 spectrum. The observations were begun on 7 April and continued until 23 July with most of the data collected in late May to July. There were a total of 240 h of data available. As described in the previous report³ the data were scaled for each observing run to have the same daily variance. The daily normalization factor used is highly correlated to the integrated signal power in the 5-min band suggesting the variations in power are due to instrumental changes. Although the analysis was done using this daily normalized data, the results were rescaled using the average normalizing factor to provide an approximate scale in m s^{-1} .

An estimate of the number of peaks in the spectrum that are significantly above the noise can be made by examining the cumulative distribution of the spectrum. A power spectrum computed from a normally distributed noise source will have an exponential distribution, thus the logarithm of the cumulative distribution will decrease linearly with power (see ref. 4). A departure from a straight line is an indication of the presence of significant spectral features and the slope of the line is a measure of the variance (σ^2) in the spectrum.

Figure 2 shows this plot for the 45–105 μHz range of the 1979 spectrum. The actual data are represented by large dots. The smaller points are for a spectrum computed the same way from data constructed by taking the original data for each day in reversed order, thus keeping the original window function. Any coherent signal present in the original data will be eliminated by this procedure. The noise statistics for the modified data power spectrum should be unchanged for periods up to the average daily observing time, that is, for frequencies higher than 40 μHz . In the actual data there are more peaks with value above $0.25 \text{ m}^2 \text{ s}^{-2}$ per 0.11 μHz than in the reversed data. The level at which the actual distribution begins to depart from the noise distribution is around 2.5σ . As the spectrum was computed with a resolution of 0.02 μHz but has a natural resolution of only 0.11 μHz , there will be about five points shown for each significant peak in the spectrum. Also, because the data have gaps at night, the day sidelobe structure introduces 2–4 apparently significantly artificial peaks for each true peak. This reduces the number of independent spectral estimates to about 200. These considerations suggest that there are about 10 independent peaks above the $2.5\text{-}\sigma$ level.

As the spectrum appears to be dominated by the sidelobe structure from the observing times, a procedure must be performed to eliminate these sidelobes. An iterative peak removal

technique was used to find and remove the peaks one at a time. First an FFT was computed and the largest peak determined. Next that peak was accurately found with a fine resolution simple Fourier transform in the vicinity of the peak. Finally the corresponding sinusoidal signal was subtracted from the original data. This procedure was repeated, producing a list of frequencies free of day sidelobes. Selecting peaks with amplitudes greater than about 30 cm s^{-1} , 14 peaks were found in the frequency band 45–105 μHz (see Table 1). Using the same amplitude criterion, another eight peaks were found at frequencies in the range extending to 165 μHz . These higher frequency peaks will not be discussed here.

To see the effect of removing these peaks from the original data, the spectrum of the residual data after subtracting these 22 sinusoids is shown in Fig. 1b. Figure 1b is shown to the same scale as Fig. 1a. Clearly, far more than 14 peaks in the band shown have been removed in the resulting spectrum. This method of peak identification appears to be a powerful tool for analysis of power spectra computed from data with complicated windows. The method finds a table of periods, phases and amplitudes that best clean the spectrum. From experiments with artificial data, we have concluded that while the correct peak is usually found, a sidelobe introduced by the window is occasionally selected. This is a weakness of the procedure that requires care in examining the resulting peaks. We have also found that the amplitudes determined from data with a complicated window are not to be relied on.

The method described above relies on the single assumption that the examined signals consist of coherent sine waves. It is, therefore, not surprising that for each frequency removed, a complete day-sidelobe set of lines disappears from the spectrum. A further check is necessary to determine whether or not the signal is coherent for the full span of the observations. For that purpose, the data with the tabulated frequencies removed were divided into two parts, 7 April–15 June and 16 June–23 July. The spectra of each part separately did not show any of the removed frequencies. If any of the frequencies removed were not present in both halves of the data, they would have been artificially put into the cleaned data by the subtraction of the sinusoid over the entire interval. In that case they would show in the spectrum of one or both halves of the cleaned data.

Two of the 14 peaks are apparently day aliases at $1/4$ and $1/8$ of a day. (With only 3 months of data the previously reported peak at 160.01 min cannot be distinguished from $1/9$ of a day and is masked by the sidelobes of the lower day-harmonics.) This leaves 12 peaks for further analysis. There is some

ambiguity in the true identification of the largest of the remaining peaks. The first 1/day sidelobe is of the same amplitude as the main peak. This could be due to two true peaks with separation 1/day or 2/day. By careful examination of the peak shapes, we have identified the peak as a true peak at 62.29 μHz with a smaller peak at 96.94 μHz , although the identification could have been made as a true peak at 73.84 μHz with smaller peaks at 50.72 and 96.94 μHz . The alternative identifications are shown in Table 1.

Interpretation

To interpret the peaks found in this spectrum, we must seek guidance from theory. Estimates for the spacing of g -mode oscillations made from the full asymptotic approximation⁵ and calculated from complete solar models (G. Berthomieu, J. Provost and J. Christensen-Dalsgaard) suggest that the g -mode oscillations of the same degree l should be about equally spaced in period for order large enough ($n > 6$). Therefore we checked the list of prominent peaks for equal spacing in period. Three of the four largest peaks are separated in period by 15.5 min. This is very close to the spacing of the standard solar model for g -modes with degree 2 and order greater than 10.

To aid in further identification of the modes we can use the asymptotic approximation⁵ which shows that for sufficiently large order n the period T is:

$$T = T_0 \left(n + \frac{l}{2} - \frac{1}{4} \right) / (l(l+1))^{1/2}$$

If we assume that the three peaks with 15.5 min separation are part of an $l=2$ series, we can calculate T_0 and then the probable spacing for the $l=1$ and $l=3$ series. Doing this we see that most of the remaining peaks are likely to be part of the $l=1$ series. If we have an estimate for l , we can plot the observed periods as $T(l(l+1))^{1/2}$ against $(k+l/2-1/4)$ where k is an integer increasing with period. If the peaks are consistent with the model and we have assigned the correct degree l , the points will all lie on a straight line. From the intercept of this line we can determine both the best value for n and check the $-1/4$ term: this has been done in Fig. 3.

The three largest peaks were used to define the $l=2$ series and thus to find T_0 . The rightmost dots in Fig. 3 represent the three largest peaks which determine the $l=2$ series. The other modes identified in Table 1 are also shown. The line was found from the three large $l=2$ peaks only; the other peaks are organized in the expected g -mode structure. The T_0 implied by the identified modes is 38.6 ± 0.5 min. The intercept of the best fit line with the abscissa is expected to be 0.25 and is found to be 0.2 ± 0.1 , thus the order n is most likely correctly determined.

Note that in several cases two peaks have been assigned to one order n in the $l=1$ series. These peaks have an average separation of 1.2 μHz and could be evidence of rotationally split modes. Assuming the $l=1$ modes are $m=\pm 1$, the 1.2- μHz separation would correspond to a rotation rate $\Omega = 7.5 \times 10^{-6} \text{ s}^{-1}$. The rotational splitting kernel for these modes has a broad maximum between 0.2 and 0.5 R . Assuming a smooth form for $\Omega(r)$ as in Fig. 2 of ref. 6, J_2 can be calculated and is of order 1.7×10^{-6} . This value differs from that found by Hill *et al.*⁷.

Now is the Stanford instrument sensitive to degree $l=2$ g -mode oscillations? For the previously reported acoustic mode oscillations in the 5-min range, the Stanford velocity differencing scheme is most sensitive to modes of degree $l=3$ to 5 but is not sensitive to modes of degree $l=1$ or $l=2$ (refs 8, 9). For low-frequency g -modes, Gough and Christensen-Dalsgaard⁹ have shown that for periods around 160 min the Stanford instrument is most sensitive to modes with degree $l=5$ to 7. For

Table 1 Frequencies of the 14 peaks found in the 45–105 μHz range of the 1979 spectrum

Frequency (μHz)	Period (min)	Power ($\text{m}^2 \text{s}^{-2}$)	l	n
58.25	286.1	0.15	1	10
59.52	280.0	0.52	1	10
64.13	259.9	0.17	1	9
65.30	255.2	0.22	1	9
73.43	227.0	0.18	1	8
73.84*	225.7	0.98	1	8
83.50	199.6	0.09	1	7
95.70	174.1	0.23	1	6
96.94	171.9	0.31	1	6
50.72*	328.6	0.07	2	20
56.33	295.9	0.49	2	18
59.52	280.0	0.52	2	17
63.12	264.1	0.42	2	16
66.65	250.0	0.13	2	15
46.23	360.5	0.25	Day/4	
62.29	267.6	0.98	—	
92.21	180.7	0.10	Day/8	

The frequencies are in μHz with the corresponding periods shown in minutes. The power is shown as $\text{m}^2 \text{s}^{-2}$ per 0.11 μHz . The classification of n and l are described in the text.

* Peaks are an alternative identification for the peak at 62.99 μHz and its associated sidelobes. Note that the peak at 59.52 μHz could be identified as part of either an $l=1$ or $l=2$ series and is included in the table twice.

longer periods however, the instrument begins to be relatively more sensitive to modes with lower degree. This variation with period, unlike the case for the p -modes, is due to the importance of the horizontal motions induced by the internal gravity modes. To estimate properly the sensitivity to the various modes, the oscillation energy must also be considered. In the simplest case, the combined effect of equipartition of energy between modes and the mode visibility with the Stanford instrument yields the greatest sensitivity for $l=1$ and $l=2$ at these longer periods.

The largest peak at 62.29 μHz appears not to be consistent with the asymptotic formula with the T_0 we have found. If the alternative choice of the 62.29–73.84 μHz pair were chosen, the largest peak would be at 73.84 μHz which would be identified as $l=1$, $n=8$. In this case a peak at 50.7 μHz shows up in the peak finding procedure and would be identified as $l=2$, $n=20$. We also note that the frequency of this peak is almost identical with the frequency difference between 5-min modes with order and degree $(n, l=1)$, $(n, l=0)$ and close to the difference between $(n, l=2)$, $(n, l=1)$. This beat frequency could show up in the Stanford observations through a nonlinear coupling in the Sun. We do not, however, find any peak near 136 μHz which would correspond to beats between acoustic modes of the same degree but with order differing by 1.

The theory predicts several hundred periods for low degree g -modes in the period interval examined. We have examined only the most visible handful of peaks and found agreement with the theory. The relative peak sizes and simplicity of the analysis leads us to the present mode identification. Alternative identifications are quite possible but they could only be tested with more data from other years or from other observatories.

We thank John Wilcox and Jorgen Christensen-Dalsgaard for useful discussions. This work was supported in part by the Office of Naval Research under contract N00014-76-C-0207, by NASA under grant NGR05-020-559 and contract NAS5-24420, by the Atmospheric Sciences Section of the NSF under grant ATM77-20580 and by the Max C. Fleischmann Foundation. Support was also received from the CNRS and from the Ministère des Relations Extérieures.

4. Knight, J. W., Schatten, K. H. & Sturrock, P. A. *Astrophys. J. Lett.* **227**, L1153 (1979).

5. Tassoul, M. *Astrophys. J. Suppl.* **43**, 469–490 (1980).

6. Gough, D. O. *Nature* **298**, 334 (1982).

7. Hill, H. A., Bos, R. & Goode, P. R. *Phys. Rev. Lett.* **49**, 1794 (1982).

8. Scherrer, P. H., Wilcox, J. M., Christensen-Dalsgaard, J. & Gough, D. O. *Sol. Phys.* **82**, 75–87 (1983).

9. Christensen-Dalsgaard, J. & Gough, D. O. *Mon. Not. R. astr. Soc.* **198**, 141 (1982).

Received 21 April; accepted 14 September 1983.

1. Scherrer, P. H., Wilcox, J. M., Kotov, V. A., Severny, A. B. & Tsap, T. T. *Nature* **277**, 635–637 (1979).

2. Severny, A. B., Kotov, V. A. & Tsap, T. T. *Nature* **259**, 87 (1976).

3. Scherrer, P. H. & Wilcox, J. M. *Sol. Phys.* **82**, 37–42 (1982).

Neodymium isotopic evidence for Galapagos hotspot–spreading centre system evolution

Surendra P. Verma^{*,†}, J.-G. Schilling & D. G. Waggoner

Graduate School of Oceanography, University of Rhode Island, Kingston, Rhode Island 02881, USA

* Instituto de Geofísica, Universidad Nacional Autónoma de México, Ciudad Universitaria, Delegación Coyoacán, D.F. 04510, Mexico

Nd and Sr isotope variations along the Galapagos spreading centre support Morgan's preferential subcrustal flow channel connecting the Galapagos mantle plume with the northeastward migrating ridge axis. Decoupling of La/Sm and Nd isotope variations near 91° W further suggests that basalts are derived from smaller degrees of melting and are deeper beneath this elevated part of the ridge.

ATTEMPTS to delineate upper mantle plume flow patterns and interaction with the large-ion depleted asthenosphere were pioneered by Vogt¹ and Schilling² for the Iceland mid-ocean ridge-centred case, using either morphological features or geochemical tracers contained in basalts erupted along the ridge axis as probes of the uppermost mantle²⁻⁴. The study of mantle plume flow patterns about plate-centred hotspots, such as Hawaii, is still out of reach because it would require drilling through the oceanic crust.

An intermediate type of hotspot–ridge configuration where mantle flow patterns may be detectable is a ridge that was once located on a plume but has since migrated, such as proposed by Morgan⁵ for the Galapagos hotspot. Considering that the plume is a source of asthenospheric material and the ridge a sink because of the accretion by cooling and crystallization, Morgan suggested that a preferential subcrustal flow channel may develop and progressively extend between the hotspot and the nearest point on the migrating ridge. The discharge of the plume material at the ridge axis would produce a geochemical anomaly and constructional volcanism preferentially deposited on the plate under which the hotspot and the flow channel are located, forming with time a single chain of islands or seamounts such as shown in Fig. 1. Morgan predicted that the discharge is now occurring just north of Darwin Island around 92° W.

Independently, Schilling and co-workers^{6,7} have shown that La/Sm variation along the Galapagos spreading centre (GSC) reaches a maximum around 91° W near the Galapagos fracture zone (FZ), whereas the subsequent determination of ⁸⁷Sr/⁸⁶Sr variation on the same samples⁸ reached a maximum around 92° W where Morgan had predicted.

The decoupling between La/Sm and ⁸⁷Sr/⁸⁶Sr maxima could be attributed to slightly smaller degrees of partial melting near 91° W as also evident petrologically^{7,9}, or by some seawater effects on basalts near 92° W, either by assimilation of seawater in the magma chamber, or assimilation of hydrothermally-altered basalts by magma stopping, or simply low-temperature alteration of the basalts since erupted.

The latter possibility was readily dismissed because no evidence was found from Fe₂O₃/FeO, major and trace element tests, thin-section inspections and leaching studies on the rocks^{8,10,11}, which would suggest such alteration.

To test the other possibilities we now report on the ¹⁴³Nd/¹⁴⁴Nd variation along the GSC profile, using a smaller selected set of 11 samples (Fig. 2). Because Nd concentration in seawater is 6 orders of magnitude smaller than that of Sr, seawater weathering or contamination by assimilation does not significantly affect the ¹⁴³Nd/¹⁴⁴Nd of the basalt or of the magma, unless the seawater/rock (or magma) mass ratio is excessively large¹²⁻¹⁵. This is true despite the ¹⁴³Nd/¹⁴⁴Nd isotopic composition of seawater being significantly lower than that of the basalts^{14,16}.

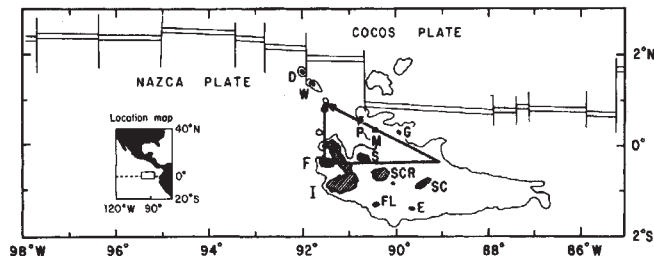


Fig. 1 Map of the Galapagos spreading centre including the Galapagos Islands (F, Fernandina; I, Isabela; M, Marchena; P, Santa Cruz; W, Wolf; D, Darwin; G, Genevosa; S, San Salvador; SC, Santa Cruz; FL, Floreana and E, Española). GSC separates the Cocos plate on the north and Nazca on the south. The arrows show for a 5-Myr total motion how the vector sum of hotspot motion (horizontal arrow) and rise crest migration (vertical arrow) relative to the Nazca plate gives the trend of M-P-W-D islands (the map is modified after Morgan⁵ and Hey³⁵).

The ¹⁴³Nd/¹⁴⁴Nd results are presented in Table 1 with their locations shown in Fig. 2. The basalts were selected so as to provide a fairly even distribution of samples along the ridge axis as well as a good representation of ⁸⁷Sr/⁸⁶Sr, Sr and rare-earth variations previously reported on the same samples⁶⁻⁸. As anticipated there is an anticorrelation between ¹⁴³Nd/¹⁴⁴Nd and ⁸⁷Sr/⁸⁶Sr distribution along the GSC (Figs 2 and 3). Geographically speaking, a minimum in ¹⁴³Nd/¹⁴⁴Nd is reached at 92° W where ⁸⁷Sr/⁸⁶Sr is maximum, and not 91° W as the La/Sm first suggested, thus giving further credibility to Morgan's channel flow model⁵.

A least-squares straight line fit (two-error regression)^{17,18} through the covariation of ¹⁴³Nd/¹⁴⁴Nd with ⁸⁷Sr/⁸⁶Sr for our 11 sample population shown in Fig. 3 gives a slope of -0.19884 ± 0.052 and an intercept of 0.65280 ± 0.037 with a linear regression correlation coefficient $r = 0.91$. Assuming that the 'bulk Earth' has a ¹⁴³Nd/¹⁴⁴Nd ratio of 0.51265 (relative to BCR-1 ¹⁴³Nd/¹⁴⁴Nd = 0.51265)¹⁹, the corresponding ⁸⁷Sr/⁸⁶Sr ratio given by extrapolation of this linear relationship would be about 0.705, thus, well within the range of 0.7045 to 0.7052 previously estimated for the 'bulk Earth'^{12,20-22}.

Seawater effects

It is now established that moderate seawater alteration of fresh ocean floor basalts by submarine weathering or hydrothermal activity in the ¹⁴³Nd/¹⁴⁴Nd versus ⁸⁷Sr/⁸⁶Sr diagram of Fig. 3 enhances the ⁸⁷Sr/⁸⁶Sr at relatively constant ¹⁴³Nd/¹⁴⁴Nd (see, for example, refs 14, 15). The length of the vector is proportional to the effective water/rock mass ratio, that is the extent of seawater interaction with the seafloor rocks or the magma chamber²³.

None of our 11 basalts from the GSC show any sign of significant isotopic overprint by seawater. We must therefore rule out the possibility that the shift between the La/Sm and ⁸⁷Sr/⁸⁶Sr maxima, or ¹⁴³Nd/¹⁴⁴Nd minimum, was due to sea-

[†] Present address: Instituto de Investigaciones Electricas, Geotermia, Apartado Postal 475, Cuernavaca, Mor. 62000, Mexico.

Table 1 Nd isotopic composition along the Galapagos spreading centre

Sample*	Location		(La/Sm) EF	Sr/Nd	Sr/Sm	Sm/Nd	⁸⁷ Sr/ ⁸⁶ Sr†	¹⁴³ Nd/ ¹⁴⁴ Nd‡
	Lat. (°N)	Long. (°W)						
VG H70	1.04	85.12	—	—	—	—	0.70277 ± 2	0.51312 ± 2
P115	0.71	85.50	—	—	—	0.31	0.70275 ± 5	0.51309 ± 3
VG A1	0.71	85.50	—	—	—	—	0.70313 ± 3	0.51312 ± 1
CTW 6D-1	0.85	86.55	0.62	5.2	13.6	0.39	0.70269 ± 3	0.51309 ± 4
TR164 5D-2	0.78	88.90	0.79	8.9	22.8	0.39	0.70276 ± 5	0.51306 ± 3
TR164 7D-1	0.86	89.99	0.93	7.7	22.8	0.34	0.70290 ± 4	0.51306 ± 3
TR164 26D-4	1.90	90.95	2.32	12.4	47.7	0.26	0.70301 ± 3	0.51303 ± 2
TR164 10D-1	2.02	91.81	1.28	11.0	37.9	0.29	0.70307 ± 4	0.51299 ± 2
TR164 12D-4	2.25	92.69	0.87	9.3	29.1	0.32	0.70292 ± 3	0.51304 ± 4
TR164 13D-1	2.42	93.34	0.98	7.9	23.1	0.34	0.70305 ± 4	0.51301 ± 2
TR164 15D-3	2.52	94.06	0.87	—	20.1	—	0.70290 ± 3	0.51305 ± 4
TR164 17D-1	2.57	94.87	0.93	6.0	19.1	0.32	0.70280 ± 8	0.51307 ± 3
VG C14	2.70	95.24	—	—	—	—	0.70282 ± 3	0.51306 ± 2
VG C34	2.70	95.24	—	—	—	—	0.70283 ± 4	0.51308 ± 2
DS D-5	2.44	95.62	0.75	—	41.5	—	0.70274 ± 3	0.51311 ± 4
TR164 20D-1	2.24	97.86	0.57	9.8	28.3	0.35	0.70262 ± 5	0.51308 ± 2

* The data for four samples (identified VG) are from ref. 27 and those for P115 from ref. 28. Nd isotopic compositions of all the other samples are obtained in the present work; other data on these samples are from refs 7 and 8.

† The ⁸⁷Sr/⁸⁶Sr ratios were previously reported by Verma and Schilling¹¹ and are normalized to E & A standard value of 0.70800.

‡ Nd isotope compositions were determined on unspiked samples using a 30.5-cm 60° mass spectrometer (VG MM30). A mixture (4:1) of HF and HNO₃ was used for decomposing the samples using Teflon beakers and clean room facilities. The evaporated samples were finally dissolved in 1 ml of 2.5 M HCl, centrifuged and loaded onto an ion-exchange column (Bio-Rad AG 50W-XB, 200–400 mesh). Rare earth elements (REE) were separated by eluting with 6 M HNO₃ following the elutions with 2.5 M HCl and 2 M HNO₃, for a better separation of REE from Ba. Nd was isolated from other REE on a second column (HDEP coated on Teflon powder) using a reverse phase chromatographic technique^{40–42}. The isobaric contribution of ¹⁴⁴Sm to ¹⁴⁴Nd is usually <10 p.p.m. using this procedure. More details on our Nd-technique can be found elsewhere⁴³. Total procedural blank was about 0.2 ng Nd. The ¹⁴³Nd/¹⁴⁴Nd ratios are normalized to ¹⁴⁶Nd/¹⁴⁴Nd = 0.72190 and are adjusted to BCR-1 ¹⁴³Nd/¹⁴⁴Nd ratio of 0.51265. The ratio for BCR-1 measured in our laboratory is 0.512566 ± 30 (2σ, n = 6). The ¹⁴³Nd/¹⁴⁴Nd ratios reported for other samples from the GSC are also adjusted to BCR-1 ¹⁴³Nd/¹⁴⁴Nd ratio of 0.51265. Finally, the errors reported on individual ⁸⁷Sr/⁸⁶Sr and ¹⁴³Nd/¹⁴⁴Nd are twice the standard error of the mean (2σ) multiplied by 10⁵.

water contamination of one kind or another. Of course, we do not imply that seawater alteration by weathering or hydrothermal activity is not necessarily taking place along the GSC. It has been directly observed with manned submersibles^{24–26}. Furthermore, of five of the dredged localities previously reported for this rift zone^{27,28}, one Fe–Ti-rich basalt (total Fe 17% as FeO) located near the 85.5° W FZ must clearly have been affected by seawater, whereas the other four plot on the same Nd versus Sr isotopic trend that we have just established for the GSC (Fig. 3).

Mixing models

Galapagos spreading centre: The fairly regular isotopic gradients about hotspots such as shown in Fig. 2 for the GSC, probably reflect some mixing process between at least two distinct types of reservoirs which themselves show a certain degree of internal variability. Questions which remain to be answered are what geological or dynamic conditions are responsible for the mixing, when and in what rheological form the mixing took place, that is whether between solids, solid–liquid mushes or melts only. For example, is mixing taking place during broad mantle convection before any melting beneath the ridge axis, or during melting or coalescence of interstitial melts into shallow magma chambers, or between melts (or mushes) derived from distinct magma bodies and mantle sources such as during subaxial pipe-like flow as suggested specifically by Vogt and DeBoer³⁰ for the GSC? Some geochemical arguments ruling out mixing of ‘melts’, as in Vogt and DeBoer’s model, have already been provided^{7,8}. We can now provide a further argument on the basis of the linear correlation observed in Fig. 3 between ¹⁴³Nd/¹⁴⁴Nd and ⁸⁷Sr/⁸⁶Sr and the degree of curvature of predicted hyperbolic curves which for mixing of two isotopically distinct reservoirs depends on a factor *k*, the concentration ratio of (Sr/Nd) in reservoir 2 (the Galapagos mantle-plume) and reservoir 1 (the source of normal mid-ocean ridge basalts, MORBs)³¹. Such mixing curves are illustrated in Fig. 5. Figure 4 shows that Sr/Nd in the basaltic melts from the GSC co-vary with Mg-value and its geographical profile along the ridge mimic that of Mg-value previously reported^{7,8}. The Sr/Nd ratios range from 2 to 17 along the ridge axis and the range is similar for the Galapagos Islands³². Partial melting and subsequent frac-

tional crystallization, particularly of feldspar which has a strong affinity for Sr, clearly have the net effects of increasing dispersion in the critical Sr/Nd ratio. Had only mixing of magmas taken place along the GSC, *k* values of 2.5–3 or even higher, that is, a noticeable degree of curvature, would be expected (see Fig. 5). This is contrary to the observation which points to a ratio of Sr/Nd between the two sources [(Sr/Nd)₂/(Sr/Nd)₁] generally closer to or equal to unity and certainly smaller than that observed in the basalts along the GSC, thus casting further doubts that the isotopic gradients observed along the GSC is caused principally by mixing of melts within the crust. It would seem that the two endmember reservoirs must have been relatively homogeneous isotopically and with not too dissimilar Sr/Nd ratios, otherwise a greater scatter in the Nd–Sr isotopic correlation would be expected due not only to the internal isotopic variation of the two endmember components, but also to the spread in Sr/Nd of the endmembers and corresponding variable *k* values this would generate. To some extent, this is also evident from the correlation of the (La/Sm) enrichment factor (EF) with ¹⁴³Nd/¹⁴⁴Nd or ⁸⁷Sr/⁸⁶Sr shown in Fig. 3 (left inserts). With the exception of basalts located near 91° W which are clearly enhanced in their La/Sm ratios, there is generally a good correlation with relatively small dispersion of the La/Sm with the Nd or Sr isotopic ratios, as would be expected for mixing between two reservoirs isotopically-distinct but not greatly dissimilar in their Nd/Sm or Sr/Sm ratios; yet again the actual variation of the Sr/Sm ratio along the ridge is greater than expected from the small scatter of points shown in Fig. 3 (insets) and that of Nd/Sm comparable (within analytical uncertainties which are poor for Nd with instrumental neutron activation analysis).

We conclude that mixing must have taken place at subcrustal depth before the increase of dispersion in the Sr/Nd and Sr/Sm ratios (to a lesser extent, Nd/Sm) as a result of fractionation processes during emplacement of magmas into the crust. The mixing is probably taking place while the material derived from the mantle plume is flowing preferentially along the channel connecting the Galapagos hotspot and the GSC, as well as along the ridge axis, east or west of 92° W. We can also argue, as previously^{7,8}, that the Galapagos mantle plume flow could not have been radial about its centre, because isotopic discontinuities

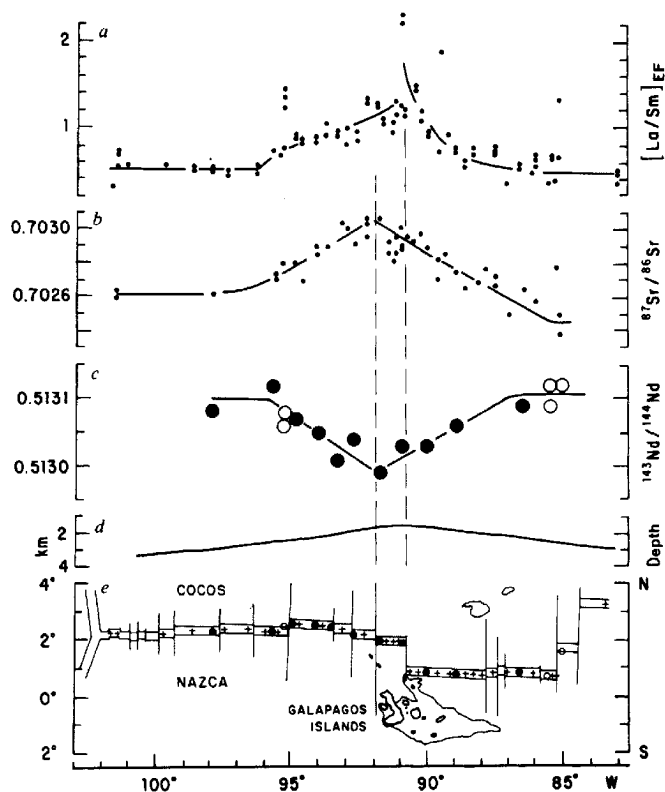


Fig. 2 Longitudinal variation of: a, (La/Sm) EF; b, $^{87}\text{Sr}/^{86}\text{Sr}$; c, $^{143}\text{Nd}/^{144}\text{Nd}$; d, water depth along ridge crest; e, a simplified map of the Galapagos area showing locations of the dredged basalts along the GSC using crosses (with La/Sm and/or $^{87}\text{Sr}/^{86}\text{Sr}$ data) and circles (with $^{143}\text{Nd}/^{144}\text{Nd}$ data). Other symbols are: small dots for (La/Sm) EF and $^{87}\text{Sr}/^{86}\text{Sr}$ data^{7,8}; large filled circles for the present $^{143}\text{Nd}/^{144}\text{Nd}$ data, large open circles for Nd isotope data from other studies^{27,28}. Note about 1° (110 km) westward shift of $^{87}\text{Sr}/^{86}\text{Sr}$ maximum (or $^{143}\text{Nd}/^{144}\text{Nd}$ minimum) relative to the (La/Sm) EF maximum.

would be expected at large fracture zones such as at 91° W, but are not observed (Fig. 2).

Galapagos Islands: In contrast, the Sr and Nd isotopic data reported for the Galapagos Islands by White and Hofmann³³, illustrated in Fig. 5, show a trend which partly overlaps that of the GSC, but extends to lower $^{143}\text{Nd}/^{144}\text{Nd}$ (or higher $^{87}\text{Sr}/^{86}\text{Sr}$). The trend shows also much greater scatter. The best linear fit through the scatter for the islands gives a slope of -0.16114 ± 0.0242 and an intercept of 0.62631 ± 0.0170 , and thus is, within the uncertainty stated, comparable with that for the GSC population previously given.

The larger scatter for the islands is not likely to be analytical judging from the two analyses of the standard BCR-1 reported by these authors. White and Hofmann³³ suggested that it was due to mixing of two magma sources or magmas which are isotopically distinct but internally variable to some degree. Subsequently White³⁴ also indicated that much of the scatter had been found to be the result of weathering effects on $^{87}\text{Sr}/^{86}\text{Sr}$ ratios. Alternatively, we suggest the possibility that the difference in the Nd-Sr isotopic scatter between the two basalt populations from the GSC and the Galapagos Islands may partly be related to distinct modes of mixing resulting from: (1) the existence of different conditions of magma segregation and emplacement into the crust beneath the islands relative to the GSC, as well as (2) the depths at which the mixing may be taking place, as the two regions differ in their tectonic and geological settings. First, we suggest that relative to the GSC, mixing beneath the islands is taking place at a shallower depth within the complex volcanic plumbing, between melts rather than solids, and these melts are petrologically more heterogeneous in composition and extent of trace element fractionation as a result of variable partial melting conditions and

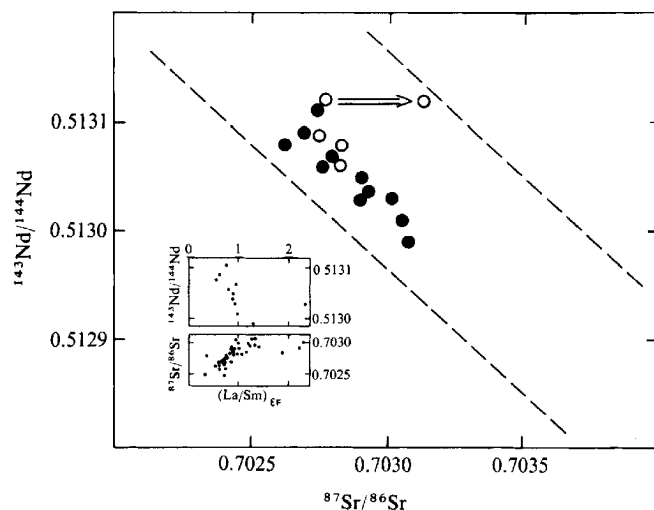


Fig. 3 Plot of $^{143}\text{Nd}/^{144}\text{Nd}$ versus $^{87}\text{Sr}/^{86}\text{Sr}$ along the GSC (large filled circles—present study; large open circles—other studies). The dashed lines show the broad 'mantle array' (drawn after Dosso and Murthy⁴⁴). The horizontal arrow shows the effects of moderate seawater alteration on the Nd and Sr isotope ratios. Insets show the plots (using small dots) of these isotope ratios versus (La/Sm) EF.

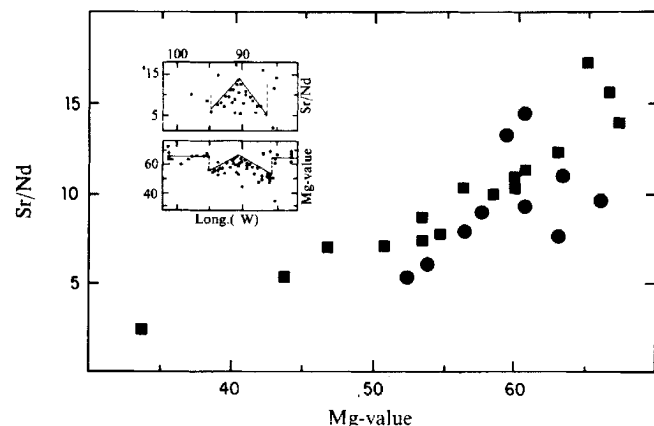


Fig. 4 Plots of Sr/Nd concentration ratio versus Mg-value ($=100 \text{ Mg}^{2+}/(\text{Mg}^{2+} + \text{Fe}^{2+})$). Large filled circles, samples from the Nd isotopes in this study; filled squares, other samples from the GSC^{7,8}. Note a positive correlation of Sr/Nd with Mg-value. Inset show longitudinal variations of Sr/Nd and Mg-value.

melt residence time within the crust. Hence, the melts from the islands have followed more complex fractional crystallization histories and larger range of degrees of partial melting (that is, with more variable Sr/Nd, Mg values, tholeiites versus alkali basalts, as it is observed on these islands^{32,33}). Note, however, that melts from the islands can still be derived from essentially two isotopically distinct reservoirs which are internally fairly homogeneous, relatively speaking. Second, because the hotspot has been intraplate for at least the past 8 Myr because of the northeastward migration of the GSC-axis³⁵, there is also the possibility that during emplacement, melts derived from the Galapagos plume have interacted with the 5–8-Myr old oceanic crust which is probably now partly weathered and isotopically affected by a seawater component resulting from hydrothermal circulation or previous submarine weathering^{36,37} and also could have been isotopically contaminated with sediments of various provenances—authigenic and detrital—the latter from the Galapagos Islands as well as from the American continent.

To illustrate our first point, Fig. 5 shows that the entire scatter observed for the islands could be bracketed by a mixing model of melts with variable Sr/Nd concentration ratios derived from two isotopically distinct mantle sources which are internally homogeneous and characterized by unique Sr and Nd isotopic

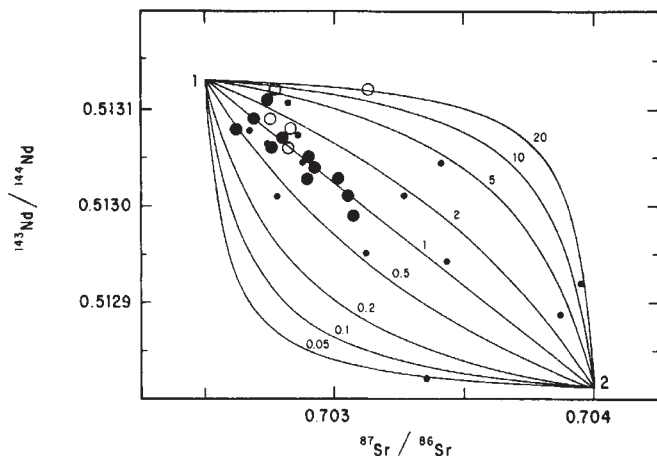


Fig. 5 Mixing model with hyperbolic mixing curves for various k values ($k = (\text{Sr}/\text{Nd})_2/(\text{Sr}/\text{Nd})_1$). In this model MORB source 1 and Galapagos plume 2 have fixed Sr and Nd isotope ratios but variable k values (curves plotted for k from 0.05 to 20). Symbols for GSC-axis are as in Fig. 3 whereas the data for Galapagos Islands³³ are plotted using smaller filled circles.

ratios. We have assumed in this first-order model that the two endmember reservoirs from which these lavas were derived and mixed are represented on one hand by normal-MORB in the region (source reservoir 1), and on the other by the Galapagos mantle-plume (source reservoir 2) chosen so as just to exceed the highest $^{87}\text{Sr}/^{86}\text{Sr}$ and the lowest $^{143}\text{Nd}/^{144}\text{Nd}$ observed by White and Hofmann³³ in their survey of the Galapagos Islands. The model shows that the Sr/Nd ratios of the endmember melts actually mixing could be randomly variable but would not in any situation have to differ from each other by more than a k factor of a maximum of 20 greater or lesser than unity (where the index of curvature $k = (\text{Sr}/\text{Nd})_2/(\text{Sr}/\text{Nd})_1$ is the concentration ratio of the two endmember melts, isotopically distinct and represented by single values representative of the endmember reservoirs 2 and 1).

Although it is difficult to weigh to what extent the scatter to the right of the linear array ($k = 1$) shown in Fig. 5 for the combined GSC and Galapagos Islands is due to mixing of 'melts' or contamination by some seawater component, a seawater component can be ruled out for the spread of values below or to the left of the ($k = 1$) linear array (or 'mantle array'). In this latter case, we must invoke mixing of melts; one endmember derived from the Galapagos plume (high $^{87}\text{Sr}/^{86}\text{Sr}$) with relatively low Sr/Nd (consequently fractionated with low Mg-value), and the other endmember, a MORB-like melt with high Sr/Nd (consequently more primary and with higher Mg-value). In support of this model is the fact that the island basalt with the lowest $^{143}\text{Nd}/^{144}\text{Nd}$ plots in Fig. 5 on the left of the linear

array ($k = 1$), which would require in the assumed mixing model a k value of about 0.05, and this is an alkali basalt with a low Mg-value from Pinta Island (Fig. 1). Remarkably, the island is close to Genovesa Island where basalts typical of MORB have recently erupted (that is, lavas which are light-REE depleted⁷).

An unlikely alternative for explaining the scatter observed on the 'left' of the 'mantle array' for the islands would be contamination of melts with some old continental-derived, granulite-dominated sediments, because such sediments would bear the Sr and Nd isotopic signature of their source which is likely to have low $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ isotopic compositions³⁸. However, there is little evidence in the sediment records of the region for such a component (for example, DSDP Holes 156–158), nor are the geology and drainage of the bordering continental mass of the western Central and South America favourable for a source of such sediments³⁹. These observations, we believe, support mixing of melts for the Galapagos Islands as indicated above.

Our interpretation is supported by the Pb isotopic data reported by White for the same Galapagos Island samples³⁴. White reports a good correlation of $^{208}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{204}\text{Pb}$ with $^{206}\text{Pb}/^{204}\text{Pb}$, with essentially no scatter (within analytical uncertainties). This is expected as in this case the denominator elements of these isotopic ratios are the same (^{204}Pb), thus the k value must be equal to unity and a straight line is expected, whatever may be the variation of the Pb concentrations of the endmember melts mixed, provided that the two reservoirs have uniform but distinct Pb isotopic compositions as called for by our model. A further test could be provided by considering the covariation of a ratio of any trace element relative to Nd with respect to $^{143}\text{Nd}/^{144}\text{Nd}$, when precise Nd data become available. In this case, a straight line would also be expected as the denominator element would be Nd and the k value would be equal to unity.

We conclude that the minimum observed at 92° W in the variation of $^{143}\text{Nd}/^{144}\text{Nd}$ along the GSC gives strong support to Morgan's preferential subcrustal flow channel connecting the Galapagos mantle plume with the northeastward migrating GSC axis. Some subaxial plume flow east and west of 92° W and dilution with the depleted asthenosphere is also suggested from the Nd and Sr isotopic gradients observed along the GSC-axis. The decoupling between La/Sm and $^{143}\text{Nd}/^{144}\text{Nd}$ in basalts erupted near 91° W most likely reflects smaller degrees of melting and segregation of melts from greater depth, a possibility which is also in harmony with petrological evidence^{7,9}.

S.P.V. thanks the Instituto de Geofísica, Universidad Nacional Autónoma de México (U.N.A.M.) for leave which made it possible to carry out the present work. We thank R. Kingsley for help in keeping the laboratory running smoothly and to A. Patino for drafting the figures. This work was supported by NSF grant OCE78-24690 and by Consejo de Estudios de Posgrado, U.N.A.M.

Received 24 June; accepted 2 September 1983.

- Vogt, P. R. *Earth planet. Sci. Lett.* **13**, 153–160 (1971).
- Schilling, J.-G. *Nature* **242**, 565–571 (1973).
- Hart, S. R., Schilling, J.-G. & Powell, J. L. *Nature phys. Sci.* **246**, 104–107 (1973).
- Sun, S.-S., Tatsumoto, M. & Schilling, J.-G. *Science* **190**, 143–147 (1975).
- Morgan, W. J. *J. geophys. Res.* **83**, 5355–5360 (1978).
- Schilling, J.-G., Anderson, R. N. & Vogt, P. *Nature* **261**, 108–113 (1976).
- Schilling, J.-G., Kingsley, R. H. & Devine, J. D. *J. geophys. Res.* **87**, 5593–5610 (1982).
- Verma, S. P. & Schilling, J.-G. *J. geophys. Res.* **87**, 10838–10856 (1982).
- Fisk, M. A., Benice, A. E. & Schilling, J.-G. *Earth planet. Sci. Lett.* **61**, 171–189 (1982).
- Hart, S. R. *Phil. Trans. R. astr. Soc. A* **268**, 573–587 (1971).
- Verma, S. P. *Chem. Geol.* **34**, 81–89 (1981).
- O'Nions, R. K., Hamilton, P. J. & Evensen, N. M. *Earth planet. Sci. Lett.* **34**, 13–22 (1977).
- O'Nions, R. K., Carter, S. R., Cohen, R. S., Evensen, N. M. & Hamilton, P. J. *Nature* **273**, 435–438 (1978).
- Piegras, D. J., Wasserburg, G. J. & Dasch, E. J. *Earth planet. Sci. Lett.* **45**, 223–236 (1979).
- Piegras, D. J. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **50**, 128–138 (1980).
- Martin, J. M., Hogdahl, O. & Philippot, J. C. *J. geophys. Res.* **81**, 3119– (1976).
- McIntyre, G. A., Brooks, C., Compston, W. & Turek, A. *J. geophys. Res.* **71**, 5459–5468 (1966).
- York, D. *Earth planet. Sci. Lett.* **5**, 320–324 (1969).
- Jacobsen, S. B. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **50**, 139–155 (1980).
- DePaolo, D. J. & Wasserburg, G. J. *Geophys. Res. Lett.* **3**, 249–252 (1976).
- Allegre, C. J., Ben Othman, D., Polve, M. & Richard, P. *Phys. Earth planet. Inter.* **19**, 293–306 (1979).
- Zindler, A., Jagoutz, E. & Goldstein, S. *Nature* **298**, 519–523 (1982).

- DePaolo, D. J. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **43**, 615–627 (1979).
- van Andel, T. J., H. & Ballard, R. D. *J. geophys. Res.* **84**, 5390–5406 (1979).
- Ballard, R. D., Holcomb, R. T. & van Andel, T. J. *J. geophys. Res.* **84**, 5407–5422 (1979).
- Crane, K. & Ballard, R. D. *J. geophys. Res.* **85**, 1443–1454 (1980).
- Cohen, R. S., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Nature* **283**, 149–153 (1980).
- Ito, E., White, W. M., von Drach, V., Hofmann, A. W. & James, D. E. *Yb. Carnegie Instn Wash.* **80**, 465–471 (1981).
- Allegre, C. J., Brévard, O., Dupré, B. & Minster, J.-F. *Phil. Trans. R. Soc. A* **297**, 447–477 (1980).
- Vogt, P. R. & DeBoer, J. *Earth planet. Sci. Lett.* **33**, 145–163 (1976).
- Vollmer, R. *Geochim. cosmochim. Acta* **40**, 283–295 (1976).
- McBirney, A. R. & Williams, H. *Mem. geol. Soc. Am.* **118**, 197p. (1969).
- White, W. M. & Hofmann, A. W. *Yb. Carnegie Instn Wash.* **77**, 596–606 (1978).
- White, W. M. *Yb. Carnegie Instn Wash.* **78**, 331–335 (1979).
- Hey, R. *Bull. geol. Soc. Am.* **88**, 1404–1420 (1977).
- Williams, D. L., Von Herzen, R. P., Sclater, J. G. & Anderson, R. N. *Geophys. J. R. astr. Soc.* **38**, 587–608 (1974).
- Humphris, S. E. & Thompson, G. *Geochim. cosmochim. Acta* **42**, 107–125 (1978).
- Carter, S. R., Evensen, N. M., Hamilton, P. J. & O'Nions, R. K. *Science* **202**, 743–747 (1978).
- Malfait, B. T. & Dinkelman, M. G. *Bull. geol. Soc. Am.* **83**, 251–272 (1972).
- Winchester, J. W. *J. Chromatogr.* **10**, 502–506 (1963).
- Schilling, J.-G. thesis, MIT, Cambridge (1966).
- Richard, P., Shimizu, N. & Allegre, C. *Earth planet. Sci. Lett.* **31**, 269–278 (1976).
- Waggoner, D. G. thesis, Univ. Rhode Island.
- Dosso, L. & Murthy, V. R. *Earth planet. Sci. Lett.* **48**, 268–276 (1980).

Induction of mammary carcinomas in rats by nitroso-methylurea involves malignant activation of *H-ras-1* locus by single point mutations

Saraswati Sukumar, Vicente Notario, Dionisio Martin-Zanca & Mariano Barbacid

Laboratory of Cellular and Molecular Biology, National Cancer Institute, Bethesda, Maryland 20205, USA

Each of nine mammary carcinomas induced by a single injection of nitroso-methylurea into 50-day-old Buf/N female rats, contained a transforming H-ras-1 gene. Molecular characterization of one of the genes revealed that the twelfth codon was GAA instead of GGA of the normal allele, encoding glutamic acid in place of glycine. These results indicate that chemical carcinogenesis represents an adequate model to study the role of transforming ras genes in human neoplasia.

DETECTION and isolation of transforming genes (oncogenes) from human tumours have provided the means to investigate, at the molecular level, events involved in the development of human neoplasia. A broad spectrum of human cancers contain such transforming genes¹, most of them identified as members of the *ras* gene family, a group of at least three structurally different genes, *H-ras*, *K-ras* and *N-ras*, which code for highly related proteins designated p21²⁻⁸. Efforts aimed at understanding the role that activated *ras* genes play in the development of human neoplasia are hampered by limitations imposed to experimental manipulation with human subjects, so we attempted to develop a well-defined animal model system. Our search for such a system was conditioned by two basic criteria: (1) a specific type of tumour must be induced under controlled experimental conditions, (2) oncogenes must be reproducibly activated in each tumour. We describe here the specific and reproducible activation of the *H-ras-1* locus in mammary carcinomas induced in Buf/N rats by a single injection of *N*-nitroso-*N*-methylurea (NMU).

Oncogenes in NMU-induced carcinomas

Gullino and co-workers have shown that exposure of 50-day-old Buf/N female rats to NMU results in the induction of mammary carcinomas in 90% of the animals with a latent period of about 60 days⁹. The carcinogenic effect of NMU appears to be hormonally regulated. Castration of the rats before NMU injection decreases incidence of tumours to negligible levels⁹. However, if the rats are ovariectomized after NMU treatment, tumours appear with the same frequency as in non-ovariectomized controls, although the number of carcinomas per animal decreases substantially and a general increase in the latency period is observed⁹. These properties made this system suitable, not only to study the reproducible activation of oncogenes, but also to investigate its putative modulation by physiological factors.

In the present studies, we have modified the experimental protocol of Gullino *et al.*⁹ by injecting a single dose of NMU (30 µg per g of body weight) to ensure induction of tumours as a result of a single carcinogenic insult. Under our experimental conditions, latency periods ranged from 6 to 12 months and resulted in the almost exclusive induction of mammary carcinomas. Mammary tumours, obtained from nine Buf/N rats which had been injected at 50 days of age with NMU (30 µg per g body weight) were minced, digested with pronase and their DNAs isolated by standard procedures¹⁰. The presence of transforming activity in these rat tumour DNAs was next investigated by transfecting NIH/3T3 mouse fibroblasts using the calcium phosphate precipitation technique^{11,12}. As summarized in Table 1, each of the nine mammary carcinoma DNAs tested induced the appearance of foci of morphologically transformed cells with efficiencies correlated well with the intactness of the tumour DNA. Maximum efficiencies were observed with DNAs

Table 1 Transforming genes in mammary carcinomas induced by nitroso-methylurea in Buf/N rats

Source of DNA	Samples tested	Transforming genes	Transformation efficiency (foci per µg DNA)
Mammary carcinomas	9	9/9	0.01–0.08
Normal breasts	10	0/10	—

Buf/N rat tumours were finely minced with the aid of razor-sharp scissors, disrupted in a loose-fitted Dounce, washed three times with phosphate-buffered saline and incubated overnight at 37 °C with 200 µg ml⁻¹ of self-digested pronase in the presence of 0.2% SDS. High-molecular weight DNA was isolated by successive extractions with 1 vol. of phenol, 1 vol. of phenol/chloroform/isoamyl-alcohol (25:24:1) and 1 vol. of chloroform/isoamyl alcohol (24:1) followed by precipitation with 2 vols of ethanol. These DNAs (20 µg per plate, 8 plates in two different transfection experiments) were used to transform NIH/3T3 cells by the calcium phosphate precipitation technique^{11,12}, as described previously⁵.

whose average molecular weight was larger than 50 kilobase pairs (kbp). In a control experiment, DNAs isolated from normal breasts of ten untreated female Buf/N rats showed no transforming activity (Table 1).

Reproducible activation of *H-ras* oncogene

Demonstration that the transformed phenotype of NIH/3T3 transfected cells was mediated by a rat oncogene and not due to spontaneous transformation required identification of donor DNA sequences stably integrated in the genome of NIH/3T3 transformants¹³. By utilizing a probe specific for rat repetitive sequences (kindly provided by Dr A. Furano) we were able to demonstrate that each of the NIH/3T3 transformants tested had taken up donor rat DNA (data not shown). We next examined whether the transforming rat sequences were related to the *ras* gene family, a group of transforming genes frequently detected in human tumours²⁻⁸. High molecular weight DNA isolated from representative NIH/3T3 transformants was digested with *Bam*HI and submitted to Southern blot analysis for detection of sequences related to the *onc* genes of the BALB and Kirsten strains of murine sarcoma viruses (MSV). BALB-MSV was originated by recombination of an ecotropic murine leukaemia virus (MuLV) with the mouse *H-ras* locus¹⁴, whereas Kirsten-MSV was generated by transduction of rat *K-ras* sequences by Kirsten-MuLV^{15,16}. In the case of the Kirsten-MSV *onc* gene probe, each of the transformants exhibited a hybridization pattern identical to that of control NIH/3T3 DNA, indicating that rat *K-ras* genes were not responsible for their malignant phenotype. In contrast, nine NIH/3T3 transformants tested, each derived from a different NMU-induced tumour, exhibited a DNA fragment not present in control mouse

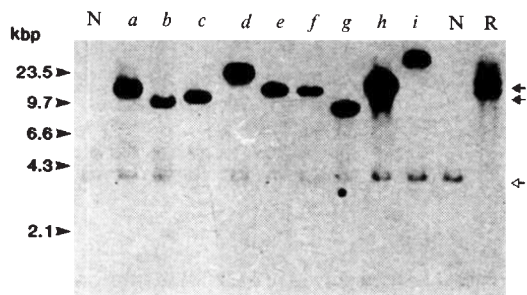


Fig. 1 Detection of H-*ras*-related sequences in NIH/3T3 mouse cells transformed by DNAs isolated from nine mammary carcinomas induced by a single injection of NMU in Buf/N rats. Twenty micrograms of DNA isolated from: N, NIH/3T3 mouse cells; a-i, NIH/3T3 transformants derived from each of the nine NMU-induced tumours; and R, normal breast of a Buf/N rat, were digested with *Bam*HI (New England Biolabs) for 1 h at 37 °C, electrophoresed (18 h at 30 V) in horizontal agarose (0.7% w/v) gels and blotted to nitrocellulose filters as described by Southern³⁵. Filters were hybridized under stringent conditions (50% formamide, 5 × SSC and 42 °C) for 44 h to 2 × 10⁷ c.p.m. of a ³²P-labelled probe obtained by nick-translation of the 675 bp *Hind*III-*Bam*HI insert of pHB-1²⁰, a DNA fragment that contains most of the coding sequences of the BALB-MSV *onc* gene. This probe detects mouse and rat H-*ras* genes with equal efficiency due to the high degree of sequence homology between these genes. The stronger hybridization signal observed in the rat sequences present in each of the NIH/3T3 transformants is due to the commonly observed amplification of transfected DNA. Hybridized blots were exposed overnight to Kodak XAR-5 film at -70 °C in the presence of intensifier screens. Co-electrophoresed DNA fragments of *Hind*III-cleaved λc1857 DNA served as size standards (labelled in kbp). Migration of the H-*ras* loci present in normal rat (←) and mouse (→) DNAs is indicated by arrows.

DNA that strongly hybridized with the *ras* oncogene of BALB-MSV (Fig. 1). These *ras*-related sequences were found to cosegregate with the malignant phenotype in additional cycles of transfection. These results demonstrate that *ras* oncogenes were activated in each of the nine NMU-induced mammary carcinomas tested in the present studies.

The rat genome contains a variety of *ras* genes^{16,17}. Even under stringent hybridization conditions, H-*ras* derived probes such as the *onc* genes of BALB and Harvey MSV detect two different loci encompassed in *Bam*HI DNA fragments of 12 and 10 kbp (see lane R in Fig. 1). These loci most likely correspond to the H-*ras*-1 gene which contains its coding sequences dispersed in four exons and to the H-*ras*-2 gene, an H-*ras*-1-derived pseudogene that has lost all intron sequences¹⁷. Thus, the different sizes of the *Bam*HI DNA fragments containing H-*ras* sequences present in each of the nine NIH/3T3 transformants tested raised the question of whether the same or distinct H-*ras* loci became activated during NMU-induced carcinogenesis.

The effect of restriction endonuclease digestion on the transforming activity of representative NIH/3T3 transformants was next examined. In all cases tested, neither *Bam*HI, *Eco*RI or *Kpn*I had any effect on the transformation efficiency of these DNAs. In contrast, both *Pvu*II and *Hind*III completely abolished their biological activity. These results suggested that the variability in size of *Bam*HI fragments seen in Fig. 1 might be due to the loss of neighbouring sequences not required for transformation, a phenomenon commonly seen in transfection experiments. Thus we analysed the DNA pattern of *ras* gene sequences present in these NIH/3T3 transformants after digestion with *Pvu*II and *Hind*III, two restriction endonucleases that must recognize cleavage sites within the transforming sequences. Figure 2 shows the results obtained with *Pvu*II. Whereas genomic rat DNA yielded four fragments of 4.3 kbp, 1.7 kbp (doublet) and 1.5 kbp, representative transformants derived from each of the nine NMU-induced tumours consistently exhibited the 4.3-kbp and 1.5-kbp *Pvu*II DNA fragments.

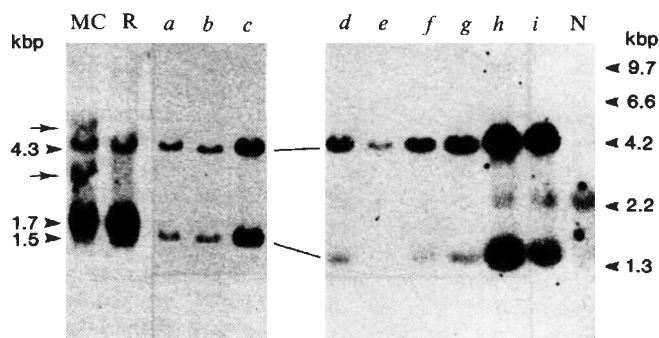


Fig. 2 The same H-*ras* oncogene is present in all NIH/3T3 transformants derived from NMU-induced mammary carcinomas. Twenty micrograms of *Pvu*II-cleaved DNAs isolated from: MC, NMU-induced mammary carcinoma 4525; R, normal Buf/N rat breast; a-i, NIH/3T3 transformants derived from each of the nine NMU-induced mammary carcinomas, and N, NIH/3T3 cells, were analysed as described in the legend to Fig. 1. Co-electrophoresed DNA size standards were those described in Fig. 1 legend. Migration of rat H-*ras* sequences including the 4.3 kbp, 1.7 kbp (doublet), and 1.5 kbp *Pvu*II DNA fragments is indicated by arrows (→). Migration of amplified H-*ras* sequences in mammary carcinoma DNA (lane MC) is also indicated (→).

Similarly, digestion of Buf/N rat normal DNA with *Hind*III generated two H-*ras*-derived fragments of 2.9 and 2.3 kbp. Only the 2.9-kbp *Hind*III DNA fragment was consistently found in all NIH/3T3 transformants (data not shown). These results conclusively demonstrate that the same H-*ras* locus has been reproducibly activated in each of the nine NMU-induced mammary carcinomas.

It should be noted that certain NMU-induced mammary carcinomas (one third of those tested) exhibited additional H-*ras* sequences (Fig. 2, lane MC) not found in normal Buf/N rat DNA (Fig. 2, lane R). The role of these amplified sequences in the carcinogenic process remains to be determined. However, they did not possess transforming activity as deduced from their absence in the multiple NIH/3T3 transformants tested throughout the present studies (Fig. 2, lanes a-i).

NMU activated oncogene an allele of H-*ras*-1

To characterize the H-*ras* oncogene activated by NMU (designated NMU-H-*ras*) we proceeded to its isolation by molecular cloning techniques. We selected a NIH/3T3 transformant, designated S2-72, which contained *ras* sequences within a single 12-kbp *Bam*HI DNA fragment. *Bam*HI digested S2-72 DNA was partially purified by sucrose gradient centrifugation, and those fractions hybridizing to the *onc* gene of BALB-MSV were ligated to *Bam*HI-cleaved DNA purified from bacteriophage λ1059¹⁸. Ligated DNAs were packaged *in vitro* into phage particles and plated onto *Escherichia coli* Q359, a bacterial strain that selects for recombinant phages that have lost the internal 14-kbp *Bam*HI DNA fragment of λ1059¹⁸. Around 10⁶ plaques were screened with the BALB-MSV *onc* gene probe¹⁹. Among the seventeen recombinants isolated, two genotypes were observed. Whereas nine phages contained the expected 12-kbp *Bam*HI DNA fragment of S2-72 DNA, eight phages contained a 6.9-kbp *Bam*HI DNA insert. However, both types of recombinant phages were able to induce malignant transformation upon transfection onto NIH/3T3 cells with similar specific activities of between 3 × 10³ and 8 × 10³ focus forming units per picomole of phage DNA. These results suggest that sequences neighbouring the NMU-activated *ras* locus were lost during cloning procedures. Because such sequences had no effect on the biological activity of the NMU-H-*ras* oncogene, we subcloned the 6.9-kbp *Bam*HI DNA fragment into pBR322. The resulting plasmid, designated pNMU-1, exhibited a specific transforming activity of 3 × 10⁴ focus forming units per picomole of DNA.

The structural features of the NMU-H-*ras* gene were studied by heteroduplex analysis between pNMU-1 and pHB-1, a

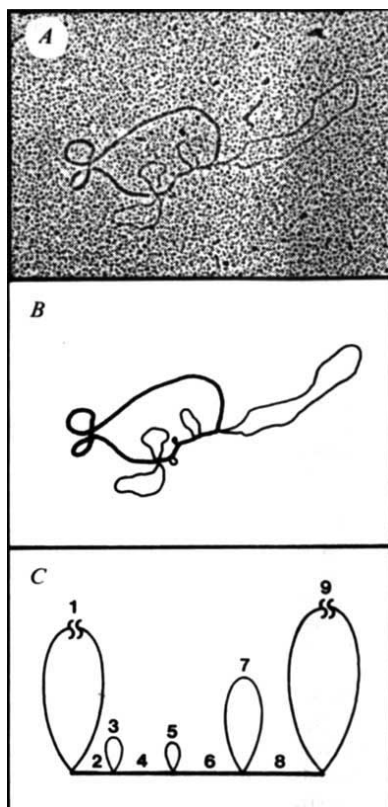


Fig. 3 Heteroduplex analysis of NMU-H-*ras* oncogene and the *onc* gene of BALB-MSV. Heteroduplexes were prepared as described previously³⁶. NMU-H-*ras* containing pNMU-1 plasmid was linearized with *Eco*RI and purified by agarose gel electrophoresis. BALB-MSV *onc* gene sequences were derived from the pHB-1 plasmid which was linearized with *Pst*I and purified as described above. **A**, A representative heteroduplex between these two DNAs. **B**, **C**, Interpretative sketches. Contour lengths (in kbp) of features are as follows: 1 = 2.41 ± 0.21 ; 2 = 0.12 ± 0.02 ; 3 = 0.16 ± 0.02 ; 4 = 0.19 ± 0.03 ; 5 = 0.14 ± 0.07 ; 6 = 0.19 ± 0.05 ; 7 = 0.74 ± 0.16 ; 8 = 0.19 ± 0.06 and 9 = 3.64 ± 0.37 . Fragments 1–9 correspond to the 6.9-kbp *Bam*HI DNA insert of pNMU-1. Fragments 2, 4, 6 and 8 correspond to the *onc* gene sequences of BALB-MSV.

plasmid containing a 675-bp *Hind*III–*Bam*HI DNA fragment of BALB-MSV which encompassed all coding sequences of this retroviral *onc* gene except the first four codons (ref. 20 and E. P. Reddy, personal communication). Figure 3 revealed four regions of homology of about 0.12, 0.19, 0.19 and 0.19 kbp in length interrupted by three single-stranded regions of 0.16, 0.14 and 0.74 kbp. These results agree with the existence in the NMU-H-*ras* gene of four exons and three introns with sizes closely matching those described for the rat H-*ras*-1 locus¹⁷. In addition, partial restriction mapping of the NMU-H-*ras* gene (Fig. 4) closely resembled that of H-*ras*-1 (ref. 17 and M. Ruta, personal communication). These results, taken together, established that the NMU-H-*ras* gene is a transforming allele of the normal rat H-*ras*-1 gene.

Point mutation in H-*ras* oncogene

The mechanism by which the rat H-*ras*-1 locus became activated in NMU-induced mammary carcinomas was next investigated. About a year ago, it was reported that a single point mutation within the coding sequences of the T24/EJ human bladder carcinoma oncogene was responsible for the acquisition of its transforming properties^{21–25}. This point mutation results in the substitution of glycine by valine as the twelfth amino acid residue of the T24/EJ oncogene-coded p21 protein^{21–25}. These observations have now been extended indicating that substitution of

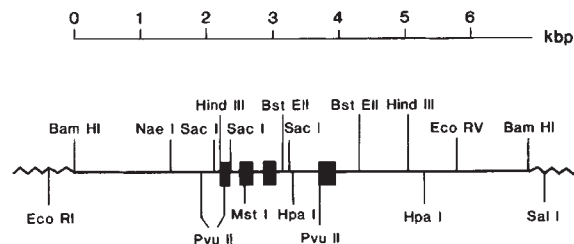


Fig. 4 Partial restriction map of the NMU-H-*ras* oncogene. The diagram depicts the location of cleavage sites for a series of restriction endonucleases within the subcloned 6.9 kbp *Bam*HI fragment of Buf/N rat DNA containing the NMU-H-*ras* oncogene. The black boxes represent exon sequences. Their location and length have been deduced from the regions of homology between pNMU-1 and pHB-1 plasmids observed in the heteroduplex analysis shown in Fig. 3. Zig-zag line represents pBR322 sequences.

glycine at position 12 by any other amino acid residue would have the same transforming consequences^{26,27}.

Thus, we proceeded to determine the nucleotide sequence of the first exon of the NMU-H-*ras* gene (Fig. 5). This sequence was then compared to that corresponding to the first exon of the normal H-*ras*-1 gene which was molecularly cloned for this purpose from normal Buf/N rat DNA utilizing λ 1059 phage as the cloning vector. First exon sequences of the NMU-H-*ras* gene and its normal allele were identified by the existence of a 52-bp *Hind*III–*Pvu*II DNA fragment that shared nucleotide sequence homology with the first exon of the T24 bladder carcinoma oncogene²⁶. The nucleotide sequence of the first exon of NMU-H-*ras* and its normal allele are depicted in Fig. 5. A single nucleotide difference consisting in the substitution of the second deoxyguanosine residue of the twelfth codon (nucleotide +35) by a deoxyadenosine, was observed. As a consequence of this point mutation, the glycine residue located at position twelve of the normal rat H-*ras*-1 p21 protein is substituted by glutamic acid in the gene product of the NMU-H-*ras* oncogene.

Substitution of a glutamine residue located at position 61 of the H-*ras* gene-coded p21 protein has recently been shown to be sufficient for the malignant activation of this locus in a human lung carcinoma cell line²⁸. Because the 61st codon of the H-*ras* gene lies within its second exon, we also determined the nucleotide sequence of the second exon of the NMU-H-*ras* oncogene and its normal allele. In this case no differences could be observed, both genes having glutamine residues at position 61. Thus, it appears that a single point mutation was responsible for the malignant activation of the H-*ras*-1 locus in NMU-induced mammary carcinomas of Buf/N rats. Interestingly, cellular p21 proteins encoded by both the normal and transforming genes contained an alanine residue at position 59 instead of the threonine residue present in Harvey-MSV p21 which is responsible for the *in vivo* phosphorylation of this retroviral protein²⁹.

The predicted amino acid sequence of the first exon of the p21 protein coded for by the Buf/N H-*ras*-1 gene is identical to that of the gene products of human H-*ras*^{21–25} and K-*ras* genes (refs 30–32 and E. Santos *et al.*, unpublished and M. Perucho, personal communications), as well as to the corresponding domains of the Harvey and BALB-MSV p21 retroviral proteins with the exception of the critical twelfth amino acid residue (ref. 29 and E. P. Reddy, personal communication). At the nucleotide level, no differences were observed between H-*ras*-1 and the *onc* gene of Harvey-MSV with the exception of a G $\rightarrow$ A transition at position +34 (Fig. 5), which is thought to be responsible for the malignant properties of this virus, although this point remains to be demonstrated. When comparisons between first exon sequences of rat and human H-*ras*-1 genes were made, only eleven (30%) third base substitutions could be detected. In contrast, little homology was observed between the first intron sequences of these two H-*ras* genes (data not shown). These findings illustrate the strong evolutionary pressure exerted to conserve the amino acid sequence of the amino-terminal domain of p21 proteins.

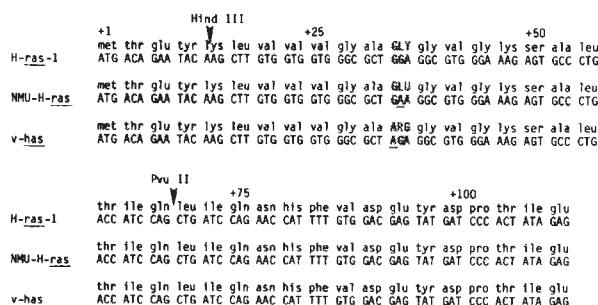


Fig. 5 Comparative nucleotide sequence analysis of the first exon of the rat NMU-H-ras oncogene and its normal allele H-ras-1. The corresponding sequence of the v-has, the *onc* gene of Harvey-MSV²⁹ is also included for comparison. Nucleotide differences between the normal H-ras-1 gene and the two oncogenes, an A instead of the normal G⁺³⁵ in NMU-H-ras and an A instead of the normal G⁺³⁴ in v-has, are underlined. These point mutations predict the incorporation of glutamic acid and arginine as the twelfth amino acid residue of transforming p21 proteins coded for by the NMU-H-ras oncogene and Harvey-MSV, respectively. The HindIII and PvuII cleavage sites that helped to localize first exon sequences are indicated. The boundaries of the first exon of H-ras-1 and NMU-H-ras were deduced by the detection of upstream terminator codons and the presence of a consensus splicing sequence at position +111. The nucleotide sequence analysis was according to the Maxam and Gilbert procedures³⁷.

NMU-induced carcinomas as model system

The present studies describe the reproducible activation of the H-ras-1 locus in an animal model system in which mammary carcinomas are specifically induced in Buf/N rats by a single dose of NMU. The specific carcinogenic action of this chemical is limited to a short period of time during which Buf/N rats reach their sexual maturity. Hormonal influence during NMU-induced carcinogenesis has been well illustrated by Gullino and co-workers by showing that ovariectomized rats have a negligible incidence of tumour formation⁹. NMU, a compound of very short half-life, was always injected intravenously to allow its rapid distribution to all the organs of the body. Therefore, its exquisite carcinogenic specificity must be mandated by a very special stage of growth and/or differentiation of certain cell type(s) of the mammary gland during sexual development. These observations suggest that understanding the mechanisms, both cellular and molecular, that lead to oncogene activation in NMU-induced mammary carcinomas should shed light on the role that cellular differentiation plays in the onset of carcinogenesis.

Induction of mammary carcinomas by NMU involves the specific activation of the H-ras-1 locus by a single point mutation. These findings validate this experimental system as an accurate model to study human carcinogenesis, at least in those human tumours in which oncogenes have been shown to become

activated by the same mechanism^{21-27,30,32}. NMU is a very potent alkylating agent that preferentially modifies deoxyguanosine residues by methylating their N⁷ and O⁶ positions (for a review, see ref. 33). As a consequence, modified deoxyguanosine residues might pair with thymidine instead of deoxycytosine, thus, leading to the generation of G→A transitions. This is precisely the mutation that appears to be responsible for the malignant activation of the NMU-H-ras oncogene characterized in the present study.

Preliminary evidence suggests that this observation may not be coincidental. Mutations affecting the second deoxyguanosine residue of the twelfth codon of H-ras-1 eliminate an *MnII* cleavage site. This restriction enzyme polymorphism was detected in all nine NMU-induced mammary carcinomas described in the present study (unpublished observations). At present we cannot discern whether the same GGA→GAA mutation identified in the one cloned NMU-H-ras oncogene is responsible for the transforming properties of the H-ras-1 oncogenes in the other eight NMU-induced carcinomas. If further investigation shows that such is always the case, it will constitute strong evidence that the oncogene activation was a direct consequence of the interaction of NMU with the DNA.

Transforming genes have also been described in mouse skin carcinomas. Balmain and Pragnell have shown the presence of an activated *ras* oncogene in each of three transplanted DMBA-induced skin carcinomas of NMRI mice³⁴. Induction of these tumours requires sequential treatment with initiator and promoter agents, thus making it more difficult to define the physiology of tumour induction. However, this experimental system offers the advantage of well-defined premalignant stages involving the appearance of multiple benign papillomas, of which only a small percentage will progress to malignancy³⁴. Whether these DMBA-induced oncogenes are also activated by mechanisms involving single point mutations remains to be determined.

The consistent presence of transforming genes in animal tumours induced by certain chemical carcinogens bears important implications on the biological significance of oncogene activation in human neoplasia. The relatively low percentage of transforming genes detected in human tumours¹ and the minute genetic alterations needed for oncogene activation^{21-28,30-32} have raised the possibility that oncogenes may not be directly involved in tumour development, but rather are a fortuitous consequence of the genetic disarray of malignant cells. However, the reproducible detection of specific transforming genes in animal model systems strongly suggests that oncogenes must have played a significant role in the development of those tumours in which they have been detected.

We thank Anne L. Arthur and Linda K. Long for technical assistance and Holly Fuhrer for photographic work. We also thank M. Sullivan (Genex Inc.) for heteroduplex analysis, A. Furano for the probe of rat repetitive sequences, E. P. Reddy and M. Ruta for communicating their results before publication and B. Zbar for encouragement and for providing animal facilities.

Received 30 August; accepted 26 October 1983.

- Cooper, G. M. *Science* **217**, 801-806 (1982).
- Der, C. J., Krontiris, T. G. & Cooper, G. M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3637-3640 (1982).
- Parada, L. F., Tabin, C. J., Shih, C. & Weinberg, R. A. *Nature* **297**, 474-478 (1982).
- Santos, E., Tronick, S. R., Aaronson, S. A., Pulciani, S., & Barbacid, M. *Nature* **298**, 343-347 (1982).
- Pulciani, S. *et al. Nature* **300**, 539-542 (1982).
- McCoy, M. S. *et al. Nature* **302**, 79-81 (1983).
- Shimizu, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 2112-2116 (1983).
- Eva, A., Tronick, S. R., Gol, R. A., Pierce, J. H. & Aaronson, S. A. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4926-4930 (1983).
- Gullino, P. M., Pettigrew, H. M. & Grantham, F. H. *J. natn. Cancer Inst.* **54**, 401-409 (1975).
- Meth. Enzym.* **12A**, (1967).
- Graham, F. L. & van der Erb, A. J. *J. Virol.* **52**, 456-463 (1973).
- Wigler, M. *et al. Cell* **11**, 223-232 (1977).
- Shih, C., Padhy, C., Murray, M. & Weinberg, R. A. *Nature* **290**, 261-264 (1981).
- Andersen, P. R. *et al. Cell* **26**, 129-134 (1981).
- Ellis, R. W. *et al. Nature* **292**, 506-511 (1981).
- Tsuchida, N. & Uesugi, S. *J. Virol.* **38**, 720-727 (1981).
- DeFeo, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3328-3332 (1981).
- Karn, J., Brenner, S., Barnett, L. & Cesareni, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5172-5176 (1980).

- Benton, W. D. & Davis, R. W. *Science* **196**, 180-181 (1977).
- Andersen, P. R., Tronick, S. R. & Aaronson, S. A. *J. Virol.* **40**, 431-439 (1981).
- Tabin, C. J. *et al. Nature* **300**, 143-149 (1982).
- Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. *Nature* **300**, 149-152 (1982).
- Taparowsky, E. *et al. Nature* **300**, 762-765 (1982).
- Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. *Nature* **302**, 33-37 (1983).
- Reddy, E. P. *Science* **220**, 1061-1063 (1983).
- Santos, E., Reddy, E. P., Pulciani, S., Feldman, R. J. & Barbacid, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4679-4683 (1983).
- Pincus, M. R., Renswoude, J., Harford, J. B., Chang, E. H. & Klausner, R. D. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5253-5257 (1983).
- Yuasa, Y. *et al. Nature* **303**, 775-779 (1983).
- Dhar, R. *et al. Science* **217**, 936-937 (1983).
- Shimizu, K. *et al. Nature* **304**, 497-500 (1983).
- McGrath, J. P. *et al. Nature* **304**, 501-506 (1983).
- Capon, D. J. *et al. Nature* **304**, 507-513 (1983).
- Margison, G. P. & O'Connor, P. J. in *Chemical Carcinogenesis and DNA Vol 1* (ed. Gover, P. L.) 111-159 (CRC Press, Boca Raton, Florida, 1979).
- Balmain, A. & Pragnell, J. B. *Nature* **303**, 72-74 (1983).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Davis, R. W., Simon, M. & Davidson, N. *Meth. Enzym.* **21**, 413-428 (1971).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).

β -Globin gene inactivation by DNA translocation in $\gamma\beta$ -thalassaemia

D. Kioussis, E. Vanin*, T. deLange†, R. A. Flavell‡ & F. G. Grosveld

Laboratory of Gene Structure and Expression, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

*Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA

†Laboratory of Biochemistry, Jan Swammerdam Institute, University of Amsterdam, PO Box 60.000, 1005 GA Amsterdam, The Netherlands

‡Biogen Inc., 14 Cambridge Center, Cambridge, Massachusetts 02142, USA

The β -globin gene present on the deletion locus in a Dutch $\gamma\beta$ -thalassaemic patient was found to be identical to the normal β -globin gene with respect to DNA sequence and its transcription in HeLa cells. DNase I sensitivity and methylation experiments show that the affected β -globin gene is present in an inactive configuration in vivo. This is the result of a translocation of a normally inactive locus next to the β -globin gene on the affected chromosome, or the deletion of sequences which are normally required for the maintenance of the active state.

THE human β -globin related polypeptide chains are encoded by a cluster of genes, located on chromosome 11 in the order 5'- ϵ - γ^G - γ^A - δ - β -3'. These genes are expressed differentially during development from the embryonic (ϵ) and fetal (γ^G , γ^A) to the adult stage (δ , β)¹. Several well-defined defects in the functioning of these genes have been described at the molecular level, which either affect the structure or the amount of β -globin protein. The latter class of defects can be subdivided into two groups, the β^+ -thalassaemias, which are characterized by a reduced amount of β -globin protein, and the β^0 -thalassaemias, in which the β -globin protein is absent in homozygotes. The defect present in several cases of β^+ -thalassaemia is caused by a point mutation, which results in the aberrant splicing of part of the β -globin precursor mRNA to a non-functional mRNA²⁻⁴. Most commonly, the defect in β^0 -thalassaemias is caused by point mutations or deletions resulting in either early termination⁵⁻⁷, or the absence of a functional mRNA. In most of these cases the γ -globin genes remain functional and active in fetal life and, in some cases, even during adult life (HPFH)^{8,9}. However, in a small group of disorders, known as $\gamma\beta$ -thalassaemias, the ϵ - and the γ -genes are also affected. These thalassaemias^{10,11} are characterized by a severe anaemia in newborn heterozygotes as a result of the reduction of the ratio of γ/α globin synthesis to about 0.5. As γ -chain synthesis ceases in the course of normal development, the disease develops into a mild β -thalassaemia with the unusual phenomenon that the HbA₂ ($\alpha_2\delta_2$) levels are normal, rather than elevated as in the classical β -thalassaemia¹⁰.

Three such cases have been reported. The first case of $\gamma\beta$ -thalassaemia¹² to be described was later shown to have a deletion of a large segment of chromosome 11, including the ϵ - γ^G - γ^A and δ -globin genes and the 5' end of the β -globin gene. In a Dutch¹¹ case, however, the entire β -globin gene and its normal 5' and 3' flanking sequences are still present, whereas the γ - and δ -globin genes are deleted. Although at different positions on the β -locus, the generation of these deletions may be the result of a common underlying mechanism, since the same DNA sequences were found to be juxtaposed to the β -gene in these unrelated cases¹³. A similar phenomenon is observed in the generation of some HPFH genotypes¹³. In a third case of $\gamma\beta$ -thalassaemia, all the globin genes have been deleted (H. H. Kazanian, personal communication).

The Dutch $\gamma\beta$ -thalassaemia is particularly interesting in that although the patients have two complete copies of the β -globin gene, they still have the haematological symptoms characteristic of a heterozygous β^0 -thalassaemia. It is surprising that the apparently normal β -globin gene in the Dutch $\gamma\beta$ -thalassaemia

is inactive in adult life. Several possible explanations have been given for this phenomenon¹¹; the thalassaemia might be the result of a double mutation, that is, the largest deletion in the ϵ - γ - δ region might be irrelevant to the β -thalassaemia, but a second, smaller deletion or a point mutation, near or in the β -globin gene, may render it inactive. Alternatively, the disease may be the result of the absence of some, or all, of the deleted DNA sequences lying far from the β -globin gene. Finally, the deletion might juxtapose the β -globin gene to a chromatin structure which is not compatible with gene expression in erythropoietic cells.

The first explanation implies that the β -globin gene region on the mutant chromosome would still be properly regulated, and that a region(s) upstream from the β -globin gene does not play a primary role in its expression. In contrast, the latter explanations imply that the upstream sequences are important in the regulation of β -globin expression. Consequently, a further analysis of this $\gamma\beta$ -thalassaemia could provide a better understanding of the mechanisms involved in the regulation of β -globin expression.

Sequence of the two allelic β -globin genes

A cosmid library was constructed using the DNA isolated from the blood of a Dutch $\gamma\beta$ -thalassaemic patient¹⁴. Four β -globin positive clones were isolated and analysed by restriction enzyme mapping. Three of the clones contained the mutant 4.2 kilobase (kb) *EcoRI* fragment, while the fourth clone contained the normal 5.2 kb *EcoRI* fragment which is the 5' end of the β -globin gene from the apparently unaffected chromosome (Fig.1). Comparison of the restriction fragments from all the clones with known restriction maps^{11,14} of the mutant and normal DNA from the patient and normal individuals established that none of the cosmid recombinants were rearranged (data not shown). The *BglII* fragments containing the β -globin gene from two of the cosmid recombinants, cosmid 5 $\gamma\beta$ from the mutant chromosome and cosmid 6N from the normal chromosome, were subcloned for sequence analysis. The three exons, the promoter regions and all the exon-intron boundaries of each of the genes were sequenced by the method of Maxam and Gilbert¹⁵. Both genes showed a sequence identical to that of the normal β -globin gene¹⁶ in these regions, that is from -100 upstream from the cap-site through the three coding segments to a position 50 base pairs (bp) downstream from the poly(A) addition site (data not shown). These data thus exclude the possibility that a point mutation in the coding sequence is responsible for the β -thalassaemic phenotype.

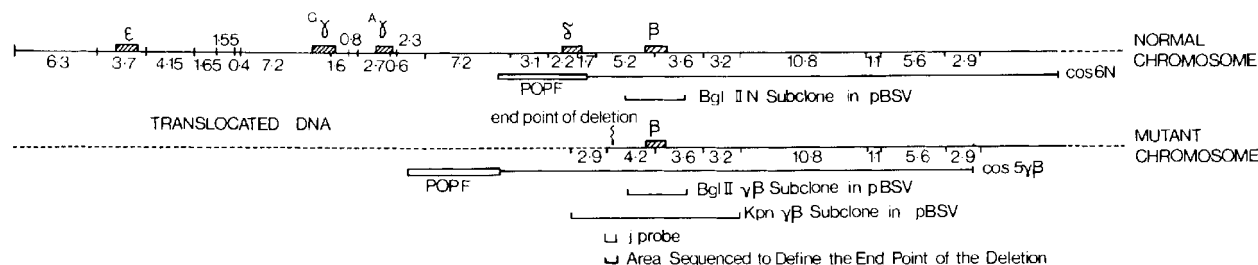


Fig. 1 Physical map of the normal and $\gamma\beta$ -thalassaemic β -globin region. Each map shows the position of the genes in relation to the *EcoRI* restriction map. The relevant clones (cos 6N and cos 5 $\gamma\beta$) were isolated from a cosmid library constructed from blood DNA of a $\gamma\beta$ -thalassaemic patient and the cosmid vector pOPF¹⁴. The brackets underneath indicate the fragments which were subcloned in the expression vector pBSV¹⁷ and the location of the junction probe (j). The arrow indicates the deletion end point in the affected locus.

Transcription of allelic β -globin genes

The fact that the DNA sequence of the affected gene was found to be identical to normal does not exclude the possibility that the gene is transcriptionally non-functional. To test the transcription of the genes, subclones in the transient expression vector pBSV¹⁷ were introduced into HeLa cells by CaPO_4 precipitation¹⁸ and the RNAs from these cells were analysed for the presence of β -globin mRNA using S_1 nuclease mapping¹⁹.

The following pBSV recombinants were constructed (Fig. 1): (1) A 4.7 kb *Bgl*II fragment from both the normal and the mutant chromosome which contains the β -globin gene and about 3 kb of flanking sequences, was cloned in the *Bam*HI site of pBSV. (2) A 14 kb *Kpn*I fragment from the mutant chromosome (from cosmid clone $\gamma\beta 4$) was cloned into the *Kpn*I site of pBSV. This fragment contains, in addition to the above sequences, flanking DNA in both 5' and 3' direction. At the 5' end of this fragment there is 2.5 kb of the flanking DNA that has been transposed next to the β -globin gene by the deletion. Insertion into the *Bam*HI or *Kpn*I site of pBSV does not affect the efficiency of transcription. A fourth plasmid containing the *Bgl*III β -globin fragment from a normal individual was used as a control.

Figure 2a shows the protected fragments after S_1 digestion, when a 5' end labelled 1,200 bp *Cvn*I probe from the 5' end of the β -gene is used. In all cases, the same amount of a 68 bp

fragment is protected, which represents the 5' end of the β -globin mRNA. A similar result is found when a 3' end labelled 700 bp *Eco*RI-*Msp*I fragment is used as a probe for the 3' end of the mRNA (Fig. 2b). In all cases, a 212 bp fragment is protected. The 5' and 3' ends of the RNA transcribed from the mutant gene are therefore indistinguishable from normal β -globin mRNA. In addition, each of the splice junctions of the mRNA was analysed as previously described³. Again, no differences were found between mature β -globin mRNA and the RNA of each of the transformants (data not shown). These results show that by these methods both the normal and the mutant gene from the patient are transcribed equally to give an mRNA indistinguishable from β -globin mRNA from reticulocytes. In addition the level of transcripts found are the same for the β -globin from the 'mutant' and normal chromosomes. Therefore, the transcription and sequence data indicate that the phenotype observed in the patient cannot be caused by a promoter defect, a 'splice' mutation, or a mutation in the coding sequence.

DNase I sensitivity of $\gamma\beta$ DNA sequences

Another explanation for the thalassaemic phenotype is the possibility that the deletion in the mutant chromosome would have brought in juxtaposition to the β -globin gene a region of the chromatin that is not expressed in erythropoietic cells, and that proximity of such a region would result in a failure to activate

Fig. 2 Transient expression of the normal and $\gamma\beta$ -thalassaemic β -globin in HeLa cells. *a*, The 5' end analysis of the subclones indicated in Fig. 1, plus two controls, a β -globin subclone from normal DNA and *in vivo* produced mRNA from reticulocytes. *b*, The 3' end analysis for the same samples. The labelled marker is Φ X174X *Rsa*I and the numbers are the length of the marker fragments in nucleotides. The lower part of the figure shows the 5' labelled 1,200-nucleotide probe and the 3' labelled 700 nucleotide probe which contain respectively 68 nucleotides from the 5' end and 212 nucleotides from the 3' end of the gene which are protected against S_1 nuclease digestion of the RNA-DNA hybrids.

Methods: The *Bgl*III and *Kpn*I restriction fragments indicated in Fig. 1 were subcloned in the expression vector pBSV¹⁷ by standard procedures. 15 μ g from each of these subclones were mixed with 25 μ g of salmon sperm DNA and used to transform half confluent 100 mm dishes of HeLa cells by the calcium phosphate method¹⁸. After 16 h the medium was changed, the cells grown for 36 h, collected and the RNA isolated by the LiCl-urea method³⁰. The β -globin present in the HeLa RNA was detected by S_1 mapping¹⁹ and end labelled DNA probes³¹.

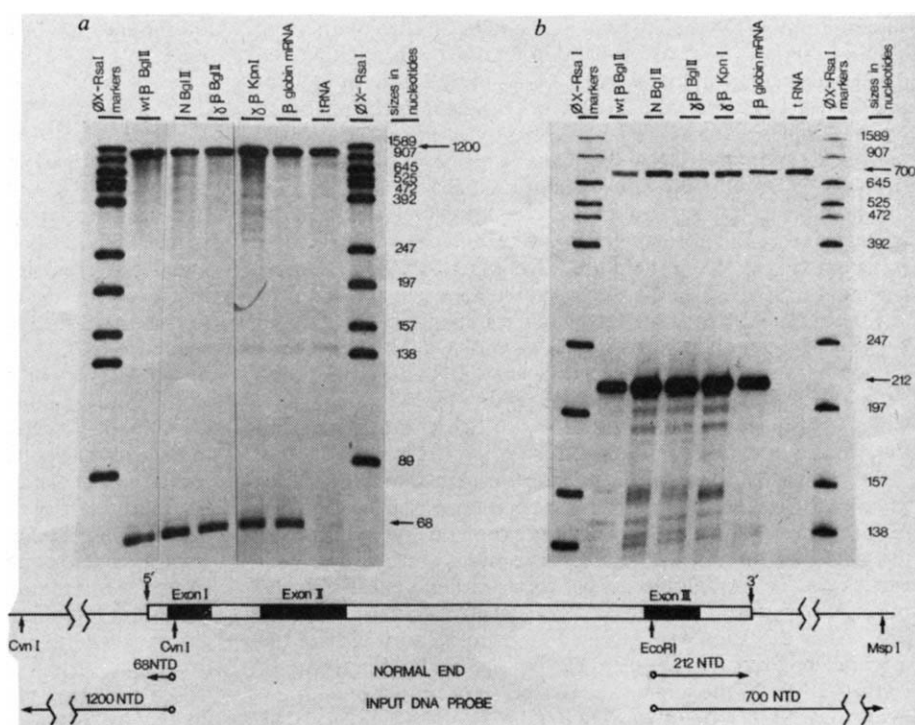


Fig. 3 DNase I sensitivity of the globin locus and the transposed sequences. *a*, Plot for 18-week fetal liver DNA; *b*, 18-week fetal brain DNA. The γ -probe hybridization is indicated by triangles, the junction-probe by squares.

Methods: Nuclei were purified from fetal brain and liver tissue as described elsewhere³² with the following modifications. The tissue was homogenized in buffer A (60 mM KCl 15 mM NaCl, 0.5 mM dithiothreitol (DTT), 0.1 mM phenylmethylsulphenyl fluoride, 15 mM Tris-HCl, pH 7.5, 0.25 M sucrose and 1 mM EDTA) using a Dounce homogenizer. The homogenate was filtered through cheesecloth and centrifuged at 5,000 r.p.m. for 10 min. The cell pellet was washed twice and lysed in 10 volumes of buffer A by the addition of Nonidet P-40 to 0.25%. Pelleted nuclei were washed twice in buffer A. The normal liver nuclei further purified on a sucrose gradient²⁷. 10^8 nuclei per ml were suspended in 60 mM KCl, 15 mM NaCl, 0.5 mM DTT, 3 mM MgCl₂, 25 mM Tris-HCl pH 7.5 and digested with DNase I (Sigma) for 1 min at 37 °C. The digestion was stopped by the addition of EDTA to 10 mM and SDS to 0.5%. The DNA was further purified by standard procedures and digested with *Eco*RI. 10 μ g of each DNA sample (different DNase I concentrations) was electrophoresed through a 0.5% agarose gel. After blotting the filters were hybridized with a 0.6 kb γ -globin fragment or a 0.5 kb junction probe (Fig. 1). The resulting bands on the autoradiographs were scanned using a Joyce-Loebl densitometer. The band intensities were normalized and plotted in relation to the non DNase I treated standard DNA sample.

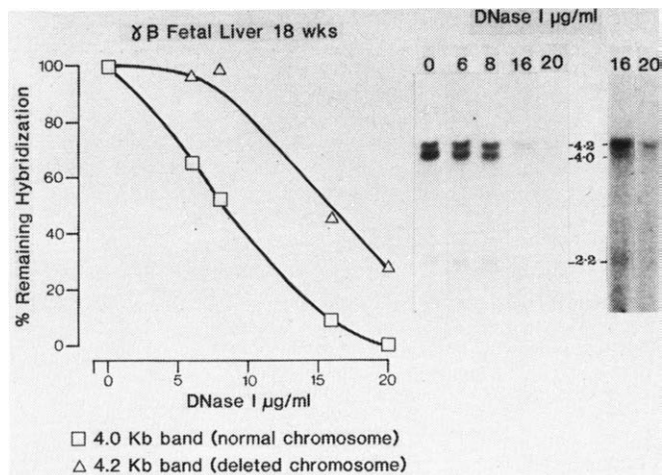
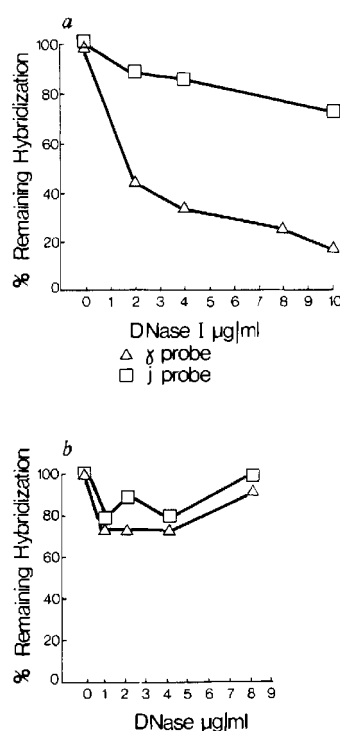


Fig. 4 The DNase I sensitivity of the normal and the mutant β -globin genes. Nuclei and DNA samples were isolated from an 18-week fetal liver of a $\gamma\beta$ -thalassaemic patient and treated essentially as described elsewhere²⁶, before double digestion with *Eco*RI and *Hin*III. The autoradiograph shows the lanes obtained with the 5' β -globin probe (Fig. 1) and an increasing concentration of DNase I. The numbers indicate the length of δ and β -DNA fragments. The plot was obtained from the blot, as described in Fig. 3. A longer exposure of the 16 and 20 μ g DNase I lanes is shown on the right.

the β -globin gene. Since transcriptional activity of a region of the genome has been associated with an increased sensitivity to DNase I²⁰ and a hypomethylation of the DNA²¹⁻²³ these assays were applied to the rearranged β -globin locus. We tested the susceptibility to DNase I of fetal liver and brain chromatin, from a normal fetus (18 weeks old) and of fetal liver chromatin from a $\gamma\beta$ -thalassaemic individual (18 weeks old). Southern blots of the DNase I treated and *Eco*RI or *Eco*RI/*Hin*III restricted DNA samples were hybridized to the following three probes: (1) An *Eco*RI-*Sph*I DNA fragment (j probe in Figs 1 and 5) from the junction point where the normal β -globin locus joins the juxtaposed DNA; this probe recognizes a 0.9 kb *Eco*RI fragment from the normal chromosome, and the 4.2 kb *Eco*RI fragment (Fig. 1) from the mutant locus. (2) A 1.8 kb *Bam*HI fragment from the 5' end of the β -globin gene which recognizes a 4.0 kb *Eco*RI/*Hin*III 5' β -globin fragment and a 2.25 kb δ -globin fragment in normal DNA, and a 4.2 kb *Eco*RI 5' β -globin fragment from the mutant locus. (3) As a control probe a 0.6 kb *Eco*RI fragment from the 3' end of the γ^A globin gene, which hybridizes to the 0.6 kb and 1.6 kb *Eco*RI fragments from the γ^A and γ^G genes, respectively.

The similarity in size of the 0.6 kb γ^A fragment and the 0.9 kb junction fragment allows a reliable comparison of the DNase I sensitivity of the two regions of the chromatin. As expected in the nuclei of the normal fetal liver samples (the tissue which expressed the fetal globin genes) the γ -globin genes from the normal liver were found to be very sensitive to DNase I (Fig. 3A). After digestion with DNase I 20 μ g ml⁻¹, only 18% of the original material remained in the 0.6 kb *Eco*RI fragment. In contrast, 80% of the junction sequences (j probe) in the normal liver were still present in the 0.9 kb *Eco*RI fragment. The

chromatin from fetal brain of the same age showed a comparable resistance to DNase I digestion for both loci (γ -globin and junction DNA, Fig. 3B). These data show that the juxtaposed (junction) DNA is resistant to DNase I in erythropoietic, and at least one non-erythropoietic tissue. If we assume that the 'inactive' configuration is transferred to the β -globin gene in the mutant locus, it predicts that the mutant locus would be less sensitive to DNase I than the normal locus in the $\gamma\beta$ -thalassaemia patient. This is confirmed when the 5' β -globin probe is hybridized to *Eco*RI/*Hin*III restricted DNA from DNase I treated fetal liver nuclei of a patient (Fig. 4). Both the 4.0 kb β -globin and the 2.25 kb δ -globin fragments from the normal locus are much more sensitive to DNase I than the 4.2 kb β -globin fragment from the mutant locus. The fact that the 4 and 2.25 kb fragments, which are smaller than the 4.2 kb fragment show a higher DNase I sensitivity excludes the possibility of differential DNase I sensitivity due to varying fragment lengths.

Methylation of $\gamma\beta$ DNA sequences

In addition to an increased DNase I sensitivity, transcriptionally active areas of the genome have been shown to exhibit hypomethylation when compared with their non-active counterparts²¹. DNA from fetal liver and brain of a normal individual, and DNA from fetal liver of a $\gamma\beta$ -thalassaemic patient were digested with *Msp*I and *Hpa*II to establish the extent of methylation of the β locus. Although both these enzymes recognize the sequence CCGG, *Hpa*II will not cleave the sequence C^mCGG or ^mC^mCGG, while *Msp*I will not cleave the sequence ^mCCGG or ^mC^mCGG²². Since most of the methylated C residues (^mC) occur in the dinucleotide ^mCG in eukaryotic DNA, the difference in the cleavage pattern of these two enzymes provides a measure for the extent of methylation in a particular region of the DNA²⁴. Southern blots of the DNA digests were hybridized to the same probes as in the DNase I experiments. Figure 5 shows that the β -probe yields the hypomethylation pattern described previously²¹ for the fetal liver, in contrast to a hypermethylation pattern in fetal brain. This is consistent with the fact that active globin genes are hypomethylated in erythroid tissues (fetal liver).

Hybridization of the β -probe to the fetal liver DNA yields a 20 kb *Hpa*II band and some high molecular weight fragments

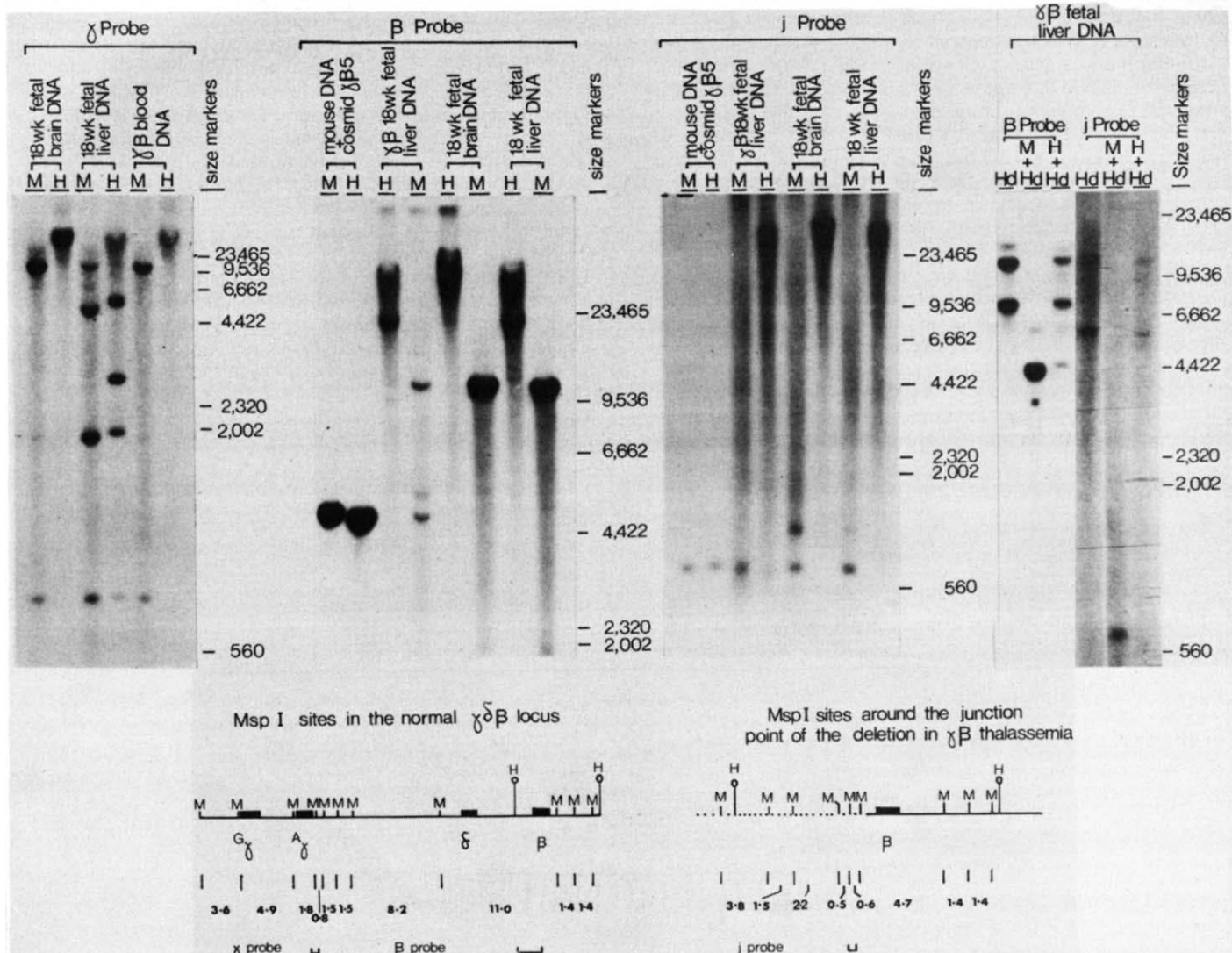


Fig. 5 Methylation of the translocated and normal $\gamma\beta$ -thalassaemia locus. DNA was isolated from 18-week fetal liver of a normal and a $\gamma\beta$ -thalassaemic patient and brain of a normal 18-week fetus. 10 μ g of DNA was digested with *Msp*I or *Hpa*II (and *Hin*III in the double digests), electrophoresed through a 0.5% agarose gel and blotted. The filters were hybridized to: a 0.6 kb 3' γ -globin fragment, a 1.8 kb 5' β -globin fragment and a 0.2 kb junction probe. The numbers indicate the fragment size of a λ X *Hind*III marker. The control lanes contained 10 μ g of mouse DNA mixed in with 1 ng of cosmid DNA.

presumably derived from the non-erythropoietic cells present in fetal liver: the 20 kb *Hpa*II band is created by cleavage of the *Hpa*II-*Msp*I sites on the 3' side of the γ^A gene and on the 3' side of the β -gene (M9 or M10 to M15 respectively in ref. 21). In fetal brain DNA, however, all of the signal has shifted to the high molecular weight range (Fig. 5, β -probe). This indicates (as previously described²¹) that the *Hpa*II sites at the 3' sites of the β -globin gene are hypomethylated in erythroid tissue. Hybridization to the fetal liver DNA of the patient (Fig. 5, β -probe) shows two complete digest bands with *Msp*I, an 11.0 kb band for the normal locus and a 4.7 kb band for the mutant locus (see maps Fig. 5). Despite the fact that all digests are controlled by an internal marker, a partial *Msp*I digest band of 5.3 kb is visible, which is probably caused by the failure of *Msp*I to cut certain C^mCGG sites²⁵. The *Hpa*II digest shows the normal 20 kb band and a high molecular weight signal, indicating that the normal locus is hypomethylated. Hybridization with the junction probe shows the expected 0.6 kb band of a complete *Msp*I digest, and a 1.1 kb partial digest band (for the same reasons as described above). The *Hpa*II digest of all the tissues examined fails to detect any low molecular weight band, indicating that all the *Msp*I sites in the $\gamma\beta$ -thalassaemia locus are methylated (Fig. 5, j probe). A *Hin*III/*Hpa*II double digest experiment (Fig. 5, $\gamma\beta$ fetal liver DNA) indeed shows that all the *Hpa*II/*Msp*I sites both at the 5' and the 3' side of

the gene (7 sites) are methylated, because the junction probe detects only the uncleaved *Hin*III fragments in the double digest. Compared with the normal locus (β probe) these results show that certainly the methylation pattern at the 3' side of the gene has been changed (there are no sites to measure in the normal locus at the 5' side of the gene). Consequently, the junction DNA is hypermethylated in the erythroid and non-erythroid tissues examined and retains this pattern (including the β -gene) when it is juxtaposed to the 5' side of the β -gene (Fig. 5). From both the DNase I sensitivity and methylation data, we conclude that the β -globin gene at the mutant locus is present in a transcriptionally inactive form.

Discussion

Van der Ploeg *et al.*¹¹ have suggested four different possibilities to explain the inactivation of the β -globin gene *in vivo* in the Dutch case of $\gamma\beta$ -thalassaemia: (1) a second mutation (for example, a stop codon or splice mutation) unrelated to the deletion, (2) inactivation of the transcriptional unit by the deletion (such as a defect in promoter sequences), (3) a *trans* effect caused by the deletion of a regulating component which affects both the normal and mutant locus and (4) a long-range *cis* effect involving either the deletion of regulatory sequences, or the transposition of inactive sequences resulting in a 'position effect'. The sequence and transcription data presented here exclude the

first two of these explanations, since both the primary sequence and the transient expression of the mutant gene are the same as for the normal gene. Our sequence data, however, do not exclude changes in the region upstream from -100, which might affect the expression of the gene *in vivo*^{26,27,33}. To this we can add that the presence of translocated sequences *per se* next to the 5' regions of the β -globin gene does not have a negative effect on the expression of the gene: the *Kpn* subclone which contains 2.5 kb of sequences from the translocated DNA is accurately and efficiently transcribed in our experimental system, although we cannot exclude that translocated sequences even further upstream might have such an effect.

The third explanation which postulates the lack of a *trans*-acting component originating in the deleted region of the chromosome in the $\gamma\beta$ -thalassaemia, leaving both β -genes active, but at a reduced efficiency, was described as unlikely¹¹, but could not be excluded. This explanation predicts that both the β -globin genes of the patient would be in a transcriptionally 'active' state. Our data clearly contradict this prediction, since they show the normal locus to be in an 'active' state, and the affected locus to be in an 'inactive' state.

Received 16 January; accepted 27 September 1983.

1. Weatherall, D. J. & Clegg, J. B. *Cell* **29**, 7-9 (1982).
2. Westaway, D. & Williamson, R. *Nucleic Acids Res.* **9**, 1777-1788 (1981).
3. Busslinger, M., Moschonas, N. & Flavell, R. A. *Cell* **27**, 289-298 (1981).
4. Spritz, R. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 2455-2459 (1981).
5. Chang, J. C. & Kan, Y. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2886-2889 (1979).
6. Moschonas, N. *et al. Nucleic Acids Res.* **9**, 4391-4407 (1981).
7. Orkin, S. H. & Goff, S. C. *J. biol. Chem.* **256**, 9782-9784 (1981).
8. Fritsch, E. F., Lawn, R. M. & Maniatis, T. *Nature* **279**, 598-603 (1979).
9. Bernards, R. & Flavell, R. A. *Nucleic Acids Res.* **8**, 1521-1534 (1980).
10. Kan, Y. W., Forget, B. G. & Nathan, D. G. *New Engl. J. Med.* **286**, 129-134 (1972).
11. van der Ploeg, L. H. T. *et al. Nature* **283**, 637-642 (1980).
12. Orkin, S. H., Goff, S. & Nathan, D. G. *J. clin. Invest.* **67**, 879-884 (1981).
13. Vanin, E., Henthorn, P., Kioussis, D., Grosfeld, F. & Smithies, O. *Cell* (in the press).
14. Grosfeld, F. G. *et al. Nucleic Acids Res.* **10**, 6715-6732 (1982).
15. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
16. Lawn, R. M., Efstratiadis, A., O'Connell, C. & Maniatis, T. *Cell* **21**, 647-651 (1980).
17. Grosfeld, G. C., deBoer, E., Shewmaker, C. K. & Flavell, R. A. *Nature* **295**, 120-126 (1982).

This observation clearly favours the fourth explanation which postulates a *cis* influence of sequences far from the β -globin gene. Whether this *cis* effect is exerted by the removal of regulating sequences, or the addition of actively suppressing sequences upstream from the β -globin gene is at present unclear. Any of these possibilities could block the normal progression of globin gene expression during normal erythropoiesis²⁸, or alter the ability of this chromosomal region to be expressed, for example, by the use of a different replication origin²⁹. Either way, in both cases, the net result is a position effect similar to those found in *Drosophila*.

The transposed DNA in $\gamma\beta$ -thalassaemia is normally found in an area of the chromatin which is DNase I insensitive and hypermethylated. Consequently, after the deletion the affected β -globin gene is present in a chromatin domain that is not expressed in erythroid tissue, resulting in the silencing of the β -globin gene.

This work was supported by the MRC and a NATO grant to R.A.F. D.K. was supported by a long-term EMBO fellowship and E.V. by NIH grant GM20069 to O. Smithies. We thank Drs M. Niermeijer and G. Grosfeld for their assistance.

18. Wigler, M. *et al. Cell* **16**, 777-785 (1979).
19. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721-732 (1977).
20. Weintraub, H. & Groudine, M. *Science* **93**, 848-858 (1976).
21. van der Ploeg, L. H. T. & Flavell, R. A. *Cell* **19**, 947-958 (1980).
22. McGhee, J. D. & Ginder, G. D. *Nature* **280**, 419-420 (1979).
23. Mandel, J. L. & Chambon, P. *Nucleic Acids Res.* **7**, 2081-2103 (1979).
24. Waaijck, C. & Flavell, R. A. *Nucleic Acids Res.* **5**, 3231-3236 (1978).
25. Busslinger, M., deBoer, E., Wright, S., Grosfeld, F. G. & Flavell, R. A. *Nucleic Acids Res.* **11**, 3559-3569 (1983).
26. Stalder, J., Groudine, M., Dodgson, J. B., Engel, J. D. & Weintraub, H. *Cell* **19**, 973-980 (1980).
27. McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D. & Felsenfeld, G. *Cell* **27**, 45-55 (1981).
28. Jones, R. W., Old, J. M., Trent, R. J., Clegg, J. B. & Weatherall, D. J. *Nucleic Acids Res.* **9**, 6813-6825 (1981).
29. Smithies, O. *J. cell. Physiol. Suppl.* **1**, 137-143 (1982).
30. Auffrey, C. & Rougeon, F. *Eur. J. Biochem.* **107**, 303-314 (1980).
31. Weaver, R. F. & Weissmann, C. *Nucleic Acids Res.* **6**, 1175-1193 (1979).
32. Wu, C., Bingham, P. M., Livak, K. J., Holmgren, R. & Elgin, S. C. R. *Cell* **16**, 797-806 (1979).
33. McKnight, S. L. *Cell* **31**, 355-365 (1982).

LETTERS TO NATURE

Can pregalactic stars or black holes generate an IR background?

B. J. Carr*†, J. McDowell* & H. Sato†

* Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

† Research Institute for Fundamental Physics, Kyoto University, Kyoto 606, Japan

Matsumoto *et al.* report¹ that they have detected an IR background in the waveband 2-5 μm which has a density $\Omega_R \sim 10^{-4} h^{-2}$ in units of the critical density ($\rho_{\text{crit}} = 5 \times 10^{-30} h^2 \text{ g cm}^{-3}$ with $H_0 = 50 h \text{ km s}^{-1} \text{ Mpc}^{-1}$) and an approximately black-body spectrum with a temperature of about 1,500 K. They claim that the intensity is too high to be explained by zodiacal light, interplanetary dust, low mass halo stars or galactic emission, and suggest that it is derived from a generation of pregalactic (population III) stars. We argue here that this is possible only if the stars began forming at a redshift exceeding 40 with a density parameter $\Omega_* \sim 1$ and a mass in the range 10^2 - $10^5 M_\odot$. With such a high density, they could avoid over-enriching the background medium with heavy elements only if they collapsed to black holes after their nuclear burning phase². These holes may also have contributed to the IR background, provided they formed optically thick accretion disks. However, we argue that holes could generate the entire background only if they have a density parameter $\Omega_B \sim 0.1$ and a mass in the range 10^6 - $10^8 M_\odot$, in which case they would be too large to have stellar precursors.

We first assume that the stars are all VMOs with a mass M exceeding $200 M_\odot$, in order to collapse to black holes^{3,4} and thereby avoid over-enrichment; in this case, their luminosity is $L = 1.3 \times 10^{38} (M/M_\odot) \text{ erg s}^{-1}$ and their surface temperature is $T_s \approx 10^5 \text{ K}$. If they all form at the same redshift z_* , their radiation will also be generated at that epoch providing their main-sequence time ($t_{\text{MS}} \approx 2 \times 10^6 \text{ yr}$) is less than the expansion time, which requires $z_* < 300 h^{-2/3}$. This means that the background radiation from the stars should just have a redshifted black-body spectrum with present temperature $T_{\text{bb}}^{\text{obs}} \approx 10^5 (1+z_*)^{-1} \text{ K}$, providing there is no absorption by grains or neutral hydrogen. The integrated energy density of the background is

$$\Omega_R = 0.004 \left(\frac{f_b X_0}{0.6} \right) \Omega_* (1+z_*)^{-1} \quad (1)$$

in units of the critical density, where the coefficient is the product of the efficiency of energy release in burning hydrogen to helium (0.007) and the fraction f_b of the hydrogen mass burnt to helium, and X_0 is the initial hydrogen abundance ($1 \geq X_0 \geq 0.75$); $f_b X_0 \approx 0.6$ for a VMO with $X_0 = 0.75$ (ref. 3). We will describe the spectrum in terms of the density parameter $\Omega_R(\nu) \equiv 4\pi i(\nu)/\rho_{\text{crit}} c^3$ (where $i(\nu)$ is the intensity per unit frequency interval) because this best emphasizes the energetic requirements of our model. The predicted spectrum is then

$$\Omega_R = 6.5 \times 10^{-4} \left(\frac{f_b X_0}{0.6} \right) \left(\frac{\Omega_*}{1+z_*} \right) \left(\frac{x^4}{e^x - 1} \right) \quad (2)$$

where $x \equiv h\nu(1+z_*)/kT_s$; this peaks at $\nu = 8.1 \times 10^{15} (1+z)^{-1} \text{ Hz}$, corresponding to $\lambda = 3.7 \times 10^{-6} (1+z_*) \text{ cm}$.

Equation (2) is compared with the data in Fig. 1. If we assume $X_0 = 0.75$, a representative fit is shown by curve *a* in Fig. 1 and

requires $z_* \approx 100$ and $\Omega_* \approx 2.5h^{-2}$. This fit does not pass through the 2- μm point; if one includes that point but neglects the 5- μm point (which is the most dubious since it could be contaminated by interplanetary dust emission), the best fit is shown by curve *b* and requires $z_* \approx 75$ and $\Omega_* \approx 1.6h^{-2}$. More generally, one might demand that the spectrum should peak between 2 and 5 μm , the range spanned by the data; the first limit is also necessary if the spectrum is to fall sufficiently to go below the optical limit⁵ at 0.5 μm . (The limit on pregalactic stars imposed by optical observations has been discussed elsewhere^{6,7}.) This corresponds to redshifts $50 < z_* < 140$, which implies that the stars certainly need to be pregalactic, the redshift of galaxy formation (z_g) usually being assumed to be about 10. The value of Ω_* required is necessarily close to 1. This may not conflict with the upper limit permitted by measurements of the cosmological deceleration parameter⁸, and it would certainly suffice to explain the dark matter problem⁹. However, it would be too large to be consistent with the observed deuterium abundance if this results from cosmological nucleosynthesis³³.

If M is allowed to fall below 200 $M_\odot$ (despite the enrichment constraints), the value required for $z_*(\propto T_s)$ decreases; the value required for $\Omega_*(\propto T_s f_b^{-1})$ first increases slightly² (as $M^{-0.2}$) and then decreases⁷ when M falls below 20 $M_\odot$. One can thus ease the constraints on Ω_* by using sufficiently small stars; furthermore, stars smaller than 4 $M_\odot$ might avoid returning any heavy elements. However, a high density of stars between 0.3 and 4 $M_\odot$ would produce a spectrum which contravenes the optical limits⁶, so this possibility does not seem very plausible. Nor does it help if the stars form over a range of redshifts or have a mass small enough that they burn out after z_* : this could extend the spectrum somewhat (including all the data points) but it does not obviate the energetic problem.

In some circumstances one might expect the spectrum to be cut off at wavelengths below the redshifted Lyman limit, $\lambda_i = 0.09(1+z_*) \mu\text{m}$, with corresponding Lyman- α emission at $\lambda_\alpha = 0.12(1+z_*) \mu\text{m}$, because of absorption by neutral hydrogen. This will not be possible if the neutral fraction is below²

$$1-x \approx 10^{-5} \Omega_g^{-1} (1+z_*)^{-3/2} \Omega^{1/2} h^{-1} \quad (3)$$

(where Ω_g is the gas density and $\Omega \geq \Omega_g$ is the total density) and sufficient ionization will be ensured by the stars themselves unless the gas clumpiness exceeds²

$$\delta \approx 5 \times 10^5 \Omega_* \Omega_g^{-2} (1+z_*)^{-3/2} \Omega^{1/2} h^{-1} \quad (4)$$

However, such a large clumpiness is not inconceivable: it could be ensured, for example, if most of the gas goes into bound clouds at decoupling or if each star is surrounded by a dense mass-loss atmosphere.

If the spectrum is cut off at λ_i , it is crucial whether the stars form over an extended redshift range. If they do not, and if $t_{\text{MS}} \ll t_*$, then the only detectable radiation will be the fraction f_p of the continuum below the Lyman limit ($f_p \approx 0.11$ for VMOs), although one of the data points may be boosted by Lyman- α emission. In this case, the observations require $0.5 \mu\text{m} < \lambda_i < 2 \mu\text{m}$, corresponding to $21 > z_* > 5$, but we still need a large value for Ω_* . Also, as shown by curve *c* in Fig. 1, which corresponds to $z_* \approx 17$ and $\Omega_* \approx 1.2h^{-2}$, the spectrum is too steep to explain all the data. On the other hand, if the stars form over an extended range of redshifts (z_1 to z_2), the Lyman- α density would dominate the background for $0.12(1+z_1) \mu\text{m} > \lambda > 0.12(1+z_2) \mu\text{m}$. If we assume that one Lyman- α photon is emitted for each ionizing photon absorbed, one can explain the observations providing $z_1 > 41$ and $z_2 < 11$. If the stars are assumed to form at a rate $\phi(z)$ in units of the critical density per redshift interval, then the Lyman- α density over the indicated wave band is

$$\begin{aligned} \Omega_\alpha(\nu) &= (0.004)\phi(z = \nu_\alpha/\nu - 1) \left(\frac{X_{\text{ofB}}}{0.6} \right) \\ &\times \left[\int_{x_1}^{\infty} \frac{x^2 x_\alpha dx}{e^x - 1} \right] / \left[\int_0^{\infty} \frac{x^3 dx}{e^x - 1} \right] \\ &= 1.2 \times 10^{-3} \phi(z(\nu)) \end{aligned} \quad (5)$$

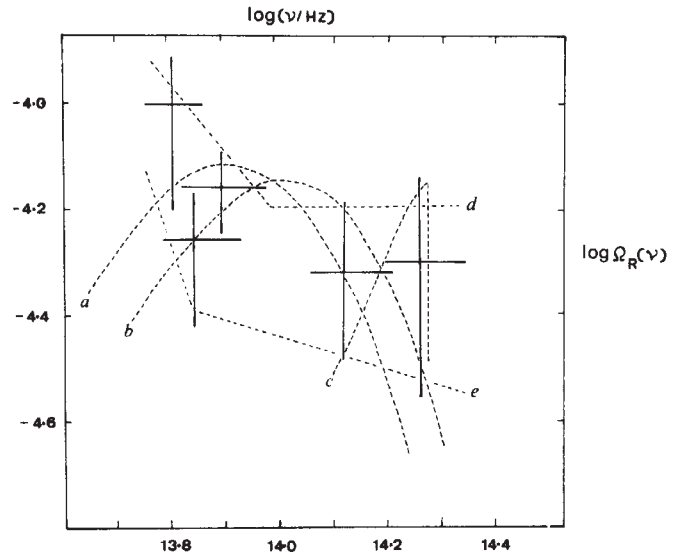


Fig. 1 Comparison of the observed IR data points, represented in terms of the spectral density parameter $\Omega_R(\nu)$ as defined in the text, with various predicted spectra. Curves *a* and *b* are the best-fit black-body spectra if one neglects the 2- and 5- μm points, respectively. These curves could represent the background generated by pregalactic starlight provided the stars form and burn at a single redshift and provided there is no absorption above the Lyman frequency. If there is a Lyman cut-off, the spectrum would be like curve *c* and would be too steep to explain all the data. If the stars form over a range of redshifts, the spectra would be more extended and could conceivably fit all the data points. This applies even if there is a Lyman cut off, since the Lyman- α line would then be spread over a wide waveband, as shown by curves *d* and *e*.

By choosing $\phi(z)$ carefully, one could in principle fit any set of data points. Curves *d* and *e* in Fig. 1 correspond to the maximum and minimum energetic requirements, the associated form for $\phi(z)$ being entirely *ad hoc*. If one allows for the extra continuum contribution, these curves require a total star density $\Omega_* \approx 1.7h^{-2}$ and $0.8h^{-2}$, respectively. Thus, providing the gas in the Universe is sufficiently clumped, Ω_* can be reduced somewhat, although it still exceeds the deuterium limit. Also, since $z_1 > 41$, star formation must still begin in the pregalactic era.

Various complications could modify the predicted spectra in Fig. 1. For example, some of the short-wavelength radiation could be absorbed by grains¹⁰⁻¹⁷; energetic requirements would then demand an even larger value for Ω_* than indicated above. Another possibility would be to assume that the stars have an extended rather than discrete mass spectrum. However, this would only be relevant if the spectrum extended below $10^2 M_\odot$ since the form of $\Omega_R(\nu)$ is independent of M for VMOs. Note also that objects larger than $10^5 M_\odot$ do not produce any appreciable starlight since such SMOs collapse due to relativistic instabilities before burning their nuclear fuel if they have no initial heavy element content¹⁸.

The black-hole remnants of pregalactic stars, as well as black holes with no stellar precursors, would inevitably accrete the background matter^{19,20}. If the holes accrete at the Bondi rate²¹, and radiate with an efficiency ϵ , each one should have a luminosity

$$L = 1.1 \times 10^{32} \epsilon M^2 n_g T_4^{-3/2} \text{ erg s}^{-1} \quad (6)$$

provided this is less than the Eddington limit ($L_{\text{ED}} = 1.3 \times 10^{38} M \text{ erg s}^{-1}$). Here M is the mass of the hole (in $M_\odot$) and n_g and T_4 are the particle density (in cm^{-3}) and temperature (in 10^4K) at the accretion radius ($R_A = 8 \times 10^{13} M T_4^{-1} \text{ cm}$). R_A is always less than the mean hole separation in our scenario. Since the background matter is heated by the radiation from accretion, the evolution of n_g and T may be very complicated (especially during the period when the holes are surrounded by individual H II regions¹⁹). However, once the Universe is ionized, we always have $n_g = 3 \times 10^{-6} \Omega_g h^2 (1+z)^3 \text{ cm}^{-3}$ on average.

The effect of the holes is crucially dependent on their radiation spectrum. If we take the standard disk accretion model, calculations²² indicate that a 'soft' spectrum of black-body form is realized if the accretion rate is less than $\dot{M}_{\text{bb}} \approx 1 \times 10^{16} M_6^{7/8} \text{ g s}^{-1}$. This is equivalent to $L < 0.02 M_6^{-1/8} L_{\text{ED}} \epsilon$, where $M_6 \equiv M/10^6 M_\odot$. In this case, the black-body temperature is

$$T_{\text{bb}} = \left(\frac{L}{4\pi R_i^2 a} \right)^{1/4} = 2.1 \times 10^4 \epsilon^{1/4} \Omega_g^{1/4} T_4^{-3/8} (1+z)^{3/4} h^{1/2} \text{ K} \quad (7)$$

where R_i is taken as $10GM/c^2$, corresponding to the inner edge of the accretion disk. We note that T_{bb} is independent of M and the spectral peak is on the low energy side of the Lyman cut-off providing $1+z < 17(\epsilon \Omega_g h^2/10^{-2})^{-1/3} T_4^{1/2}$. In this case, in trying to fit the data, the issue of whether condition (3) is satisfied is not crucial. Of course, the spectrum will not be exactly black body: it will be smeared out because of the range in M and z , and even the spectrum from an individual hole will not be perfect black body because there will be contributions from cooler points of the accretion disk with $R \gg R_i$.

The present energy density and temperature of the black-body radiation generated by the holes at redshift z are respectively

$$\Omega_{\text{IR}} = \frac{L t n_{\text{B}}}{(1+z) \rho_{\text{crit}} c^2} = 8 \times 10^{-5} \epsilon M_6 \Omega_B \Omega_g^{-1/2} (1+z)^{1/2} T_4^{-3/2} h \quad (8)$$

and

$$T_{\text{bb}}^{\text{obs}} = 2.1 \times 10^4 \{ \epsilon \Omega_g h^2 (1+z)^{-1} \}^{1/4} T_4^{-3/8} \text{ K} \quad (9)$$

Here $t = 4 \times 10^{17} \Omega_g^{-1/2} h^{-1} (1+z)^{-3/2} \text{ s}$ is the age of the Universe at redshift z , and $n_{\text{B}} = 2.5 \times 10^{-69} \Omega_B M_6^{-1} h^2 \text{ cm}^{-3}$ is the present number density of the holes. The nice point is that $T_{\text{bb}}^{\text{obs}}$ is in the range 10^3 – 10^4 K for reasonable values of ϵ , Ω_g and T_4 , and depends on z only weakly.

We now estimate the epoch z_{bb} after which the generated spectrum is approximately black body. This requires that the accretion rate be less than $\dot{M}_{\text{bb}}$, implying

$$1+z_{\text{bb}} = 17 M_6^{-3/8} \Omega_g^{-1/3} h^{-2/3} T_4^{1/2} \quad (10)$$

On the other hand, the matter temperature T_4 is determined by the balance between heating and cooling. The heating will be due to photoionizations from the fraction $(1-f_p)$ of the total energy flux above the Lyman limit, while the cooling will be primarily due to Compton scattering of electrons off the microwave background and line emission. For the range of values of f_p anticipated, one expects T to be boosted to a value in the range 10^4 – 10^5 K , with the Universe being reionized, but not appreciably higher. Since T_4 remains nearly constant, equation (8) implies that the main contribution to the background radiation is derived from z_{bb} . The contribution from $z < z_{\text{bb}}$ merely extends the spectrum to slightly higher frequencies. Before z_{bb} , the same holes might have generated X rays with a luminosity close to L_{ED} . In this case, a larger fraction of their energy would have gone into heating the background because of their Compton heating effect¹⁹ and T could be boosted well above 10^4 K in some period.

Putting $z = z_{\text{bb}}$ in equation (8), we estimate the hole mass required as

$$M = 4 \times 10^6 \left(\frac{\epsilon}{0.1} \right)^{-1.2} \Omega_B^{-1.2} \Omega_g^{0.6} \Omega_g^{-1.0} h^{-3.2} T_4^{1.5} \times \left(\frac{\Omega_{\text{IR}}}{10^{-4} h^{-2}} \right)^{1.2} M_\odot \quad (11)$$

Thus (for given values of ϵ , Ω and Ω_g) M , Ω_B and z_{bb} are determined once any one of them is specified. We note that the spectrum goes soft before galaxy formation provided

$$M < 3 \times 10^7 \left(\frac{\Omega_g}{0.1} \right)^{-0.9} \left(\frac{1+z_g}{10} \right)^{-2.7} h^{-1.8} T_4^{1.3} M_\odot \quad (12)$$

which implies

$$\Omega_B > 0.2 \left(\frac{\epsilon}{0.1} \right)^{-1} \left(\frac{\Omega_g}{0.1} \right)^{-0.1} \left(\frac{1+z_g}{10} \right)^{2.2} T_4^{0.2} \Omega^{0.5} h^{-1.2} \times \left(\frac{\Omega_{\text{IR}}}{10^{-4} h^{-2}} \right) \quad (13)$$

On the other hand, we require M to exceed the value given by equation (11) with $\Omega_B = 1$. Thus, the mass and density of the holes in this scenario can be specified fairly precisely: M has to be roughly 10^6 – $10^8 M_\odot$, corresponding to precursors in the SMO range, and Ω_B has to exceed about 0.1, which would be sufficient to provide the dark matter⁹. Equations (10) and (11) also imply that the radiation comes from a relatively recent epoch, even though the holes themselves may form much earlier. If $z = z_{\text{bb}}$ is placed in equation (9), a temperature

$$T_{\text{bb}}^{\text{obs}} = 4 \times 10^3 M_6^{0.1} \left(\frac{\epsilon}{0.1} \right)^{1/4} \left(\frac{\Omega_g}{0.1} \right)^{1/3} T_4^{-1/2} h^{2/3} \text{ K} \quad (14)$$

is obtained. Note that the temperature would be smaller if the holes had thick accretion disks, thus increasing the value of R_i in equation (7). Since both the parameters T_4 and R_i are uncertain, one cannot predict $T_{\text{bb}}^{\text{obs}}$ very precisely.

It is interesting to consider what happens if M is outside the required range. The pregalactic Universe has two spectral windows where it is optically thin²³: one in the X-ray band and one below the Lyman limit. If M exceeds the value specified by equation (12), the holes will never have a soft spectrum in the pregalactic era and will therefore contribute only to the X-ray window. For smaller masses, one can envisage the holes contributing to both windows successively, although their contribution to the IR background will be too small to explain the data if M is less than the value specified by equation (11) with $\Omega_B = 1$. It is therefore natural to suggest that the X-ray background was generated by very massive holes^{24–27}, while the IR background was generated by somewhat less massive ones.

The black holes would also accrete after galaxy formation but their luminosity would then depend on their environment²⁸. If they resided in galactic halos or clusters, their luminosity would be too low to explain the IR background, although holes traversing the disk might present detectable individual IR sources²⁹. Holes in the cosmological background could not contribute appreciably to the IR background because, even if there is a large gas density there, its temperature³⁰ must exceed 10^6 K , suppressing the luminosity by at least $\sim 10^3$. Giant holes in galactic nuclei might, however, generate an appreciable IR background³¹.

In conclusion, either stars or black holes could explain the IR background. If stars are responsible, they almost certainly have to be pregalactic VMOs with a density close to critical. If holes are responsible, they would have to be SMOs and their density and the redshift from which their radiation is derived could be somewhat smaller. Various observational features may decide which explanation is better. In particular, the detection of any anisotropy on the scale of galaxies and clusters ($\sim 1'$) or the identification of a Lyman- α line would be important. Even if the claim of Matsumoto *et al.* is not confirmed, their results nevertheless confirm previous indications³² that Ω_{IR} cannot exceed about 10^{-4} . Thus Fig. 1 and equation (11) necessarily impose interesting constraints on Ω_* and Ω_B .

Another important issue is whether the stars or black holes could contribute to the 3 K background^{10–17}. Since one of the reasons for suggesting this is that one would expect these objects to generate $\Omega_R \sim 10^{-4}$ if they provide the dark matter², the discovery of another waveband with this density may actually make this less plausible. Nevertheless, it is interesting to enquire whether pregalactic objects could generate two backgrounds with $\Omega_R \sim 10^{-4}$. Objects larger than $10^5 M_\odot$ could generate radiation efficiently only in their black-hole stage; they might generate the 3 K background at $z > 10^3$, if one invokes free-free thermalization¹³, as well as the X-ray background before z_{bb}

or the IR background after z_{bb} . Objects smaller than $10^5 M_\odot$ could contribute to the 3 K background during their nuclear-burning stage provided their light can be thermalized by grains¹⁰⁻¹⁷. They might conceivably generate both the 3 K and IR backgrounds sequentially if the grains thermalize only at sufficiently large redshifts²⁷. However, the accretion-generated radiation from their remnants would be too small to explain either background.

We thank J. R. Bond for pointing out an error in an earlier version of this paper and for subsequent helpful remarks and T. Matsumoto and S. Hayakawa for stimulating discussions. B.J.C. thanks the Research Institute for Fundamental Physics for hospitality and financial support received during this work. J.C.M. acknowledges an SERC studentship.

Received 3 July; accepted 6 October 1983.

1. Matsumoto, T., Akiba, M. & Murakami, H. *COSPAR Journal: Advances in Space Research* (eds Bignami, G. F. & Sunyaev, R. A.) (Pergamon, Oxford, in the press).
2. Carr, B. J., Bond, J. R. & Arnett, W. D. *Astrophys. J.* (in the press).
3. Bond, J. R., Arnett, W. D. & Carr, B. J. *Astrophys. J.* (in the press).
4. Ober, W. W., El Eid, M. F. & Fricke, K. J. *Astr. Astrophys.* **119**, 61 (1983).
5. Dube, R. R., Wicks, W. C. & Wilkinson, D. T. *Astrophys. J.* **232**, 333 (1979).
6. Peebles, P. J. E. & Partridge, R. B. *Astrophys. J.* **148**, 713 (1967).
7. Thorstensen, J. R. & Partridge, R. B. *Astrophys. J.* **200**, 527 (1975).
8. Sandage, A. R. *Astrophys. J.* **178**, 1 (1972).
9. Faber, S. M. & Gallagher, J. S. A. *Rev. Astr. Astrophys.* **17**, 135 (1979).
10. Rowan-Robinson, M., Negroponte, J. & Silk, J. *Nature* **281**, 635 (1979).
11. Puget, J. L. & Heyvaerts, J. *Astr. Astrophys.* **83**, L10 (1980).
12. Rees, M. J. *Nature* **275**, 35 (1978).
13. Carr, B. J. *Mon. Not. R. astr. Soc.* **195**, 669 (1981).
14. Wright, E. L. *Astrophys. J.* **255**, 401 (1982).
15. Allven, H. & Mendis, A. *Nature* **266**, 699 (1977).
16. Wickramasinghe, N. C. *et al. Astrophys. Space Sci.* **35**, L9 (1975).
17. Rana, N. C. *Mon. Not. R. astr. Soc.* **197**, 1125 (1981).
18. Fricke, K. J. *Astrophys. J.* **183**, 941 (1973).
19. Carr, B. J. *Mon. Not. R. astr. Soc.* **194**, 639 (1981).
20. Mészáros, P. *Astr. Astrophys.* **38**, 5 (1975).
21. Bondi, H. *Mon. Not. R. astr. Soc.* **112**, 195 (1952).
22. Eardley, D. M., Lightman, A. P., Payne, D. G. & Shapiro, S. L. *Astrophys. J.* **224**, 53 (1978).
23. Sato, H. *Prog. theor. Phys.* **40**, 781 (1968).
24. Carr, B. J. *Nature* **284**, 326 (1980).
25. Boldt, E. & Leiter, D. *Nature* **290**, 483 (1981).
26. Liang, E. P. T. *Nature* **286**, 39 (1980).
27. Hayakawa, S. *Proc. Semin. Fundamental Studies: Present Attitude and Future Prospect* (Colombo, Sri Lanka, 1982).
28. Carr, B. J. *Mon. Not. R. astr. Soc.* **189**, 123 (1979).
29. Iper, J. R. & Price, R. H. *Astrophys. J.* **216**, 548 (1977).
30. Gunn, J. E. & Peterson, B. A. *Astrophys. J.* **142**, 1633 (1965).
31. Takahara, F. Workshop at Nagoya University (March 7-8, 1983).
32. Hofmann, W. & Lemke, D. *Astr. Astrophys.* **68**, 389 (1979).
33. Yang, J., Schramm, D. N., Steigman, G. & Rood, R. T. *Astrophys. J.* **277**, 697 (1979).

A new limit on the nature of the galactic missing mass

Gerard Gilmore

Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

Paul Hewett

Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

It has been widely accepted that large amounts of 'missing' mass must be present in spiral galaxies to explain both the small-scale and large-scale kinematic observations¹. The small-scale data (velocity dispersions in the solar vicinity) refer to the volume within a few hundred parsecs of the Sun, and indicate that 25-50% of the gravitating mass has no detected luminous counterpart²⁻⁴. The large-scale data (velocity dispersions and rotation curves) show that this non-luminous component becomes increasingly important with increasing galactocentric distance, and is the dominant mass in the outer regions of galaxies, and in total. It is usually assumed that both the local and global missing mass are the same substance although no plausible candidate for this substance is known. The most widely assumed, and most conservative, candidate is a large population of very low-luminosity M dwarfs. These stars are

attractive, both because they exist in large numbers, and because their mass (M) to light (L) ratios are similar to those needed to explain flat rotation curves ($M/L \sim 100-1,000$). Finally, they are of very low intrinsic luminosity, and it is therefore observationally very difficult to determine their true space density from blue or visual surveys. We present here new observational results which show that low-luminosity M dwarfs are unlikely to provide the large-scale missing mass, and cannot provide the small-scale missing mass.

Low-mass luminous stars are very red, and may be readily detected and studied using current red-sensitive photographic emulsions and charged coupled device (CCD) detectors. We are therefore carrying out a survey using red-sensitive plates from the UK Schmidt Telescope in Australia which are scanned with the COSMOS machine at the Royal Observatory of Edinburgh to discover complete samples of red dwarfs. The stars are then studied individually in the near IR and/or spectroscopically to determine their space density and physical properties. We have presented elsewhere the first results (refs 5-7 and work in preparation), which are a survey of ~ 18 square degrees near the south galactic pole to $l = 17$. This is the first volume-complete determination of the local space density of low-mass stars which is complete for all stars sufficiently massive to burn hydrogen, and shows that M dwarfs do not provide the solar neighbourhood missing mass. However, very few stars with $M_v \geq +15$, mass $\leq 0.1 M_\odot$, were included in that survey. The statistical significance of our mass limit on the lowest mass luminous stars was therefore capable of improvement.

We have now extended the survey to constrain the M dwarf contribution to the larger-scale missing mass. To do this we have derived a sample of all those stars which have $l \leq 18.25$ and $B_J \geq 22$, where B_J refers to the IIIa-J+GG395 photographic passband, in a total of 50 square degrees distributed over several fields. (A discussion of the data for the other stars in these fields will be presented elsewhere.) This sample has a surface density very near 1.0 per square degree, and is restricted, allowing for the colour term in the B_J system, to stars with $M_v \geq +15$ within about 200 pc of the Sun. To determine the absolute magnitudes of the sample, and hence the space density and total mass density, we observed a random subsample of these stars (complete observations were prevented by poor weather) with the Cerro Tololo Interamerican Observatory (CTIO) 4-m telescope, RC spectrograph and CCD detector in July 1983. Spectra with signal-to-noise ratio ~ 100 were obtained in ~ 30 min per star, and are shown in Fig. 1. Also shown are spectra obtained at the same time of several low-luminosity stars with reliably determined absolute magnitudes. Comparison of the spectra clearly shows that all the programme stars have absolute magnitudes between those of GJ1002 ($M_v = +15.5$, ref. 8) and vB10 ($M_v = +18.5$, ref. 9). No star is significantly fainter than vB10, although one star for which we do not yet have a spectrum has redder optical colours. In the survey reported in refs 5 and 6, which was restricted to a volume a factor of ~ 30 smaller than the present survey, we also found just one star of very low luminosity (RG0050.5-2722, $M_v \sim +19$, refs 10, 11). An increase of a factor of about 30 has, therefore, not sampled any deeper into the luminosity function, implying either that those stars detected to date are the faintest which exist, or that the space density of objects less luminous than $M_v \sim +19$ is substantially below that of more luminous stars. In either case we may conclude that low-mass luminous stars do not contribute significantly to the total mass density in the solar neighbourhood, and are not a viable candidate for the local missing mass. Further discussion of the local mass function is given in refs 5 and 6.

The second result of this survey is that the surface density of stars with $M_v \geq +16$ and $l \leq 18$ is near 1.0 ± 0.5 per square degree, with the uncertainty being a liberal estimate based on the uncertainties in our magnitude calibration and photometry at these magnitudes. This surface density corresponds to a space density of $\leq 10^{-2} \text{ pc}^{-3}$ at the Sun. We may compare this with the values necessary to provide the local or global missing mass.

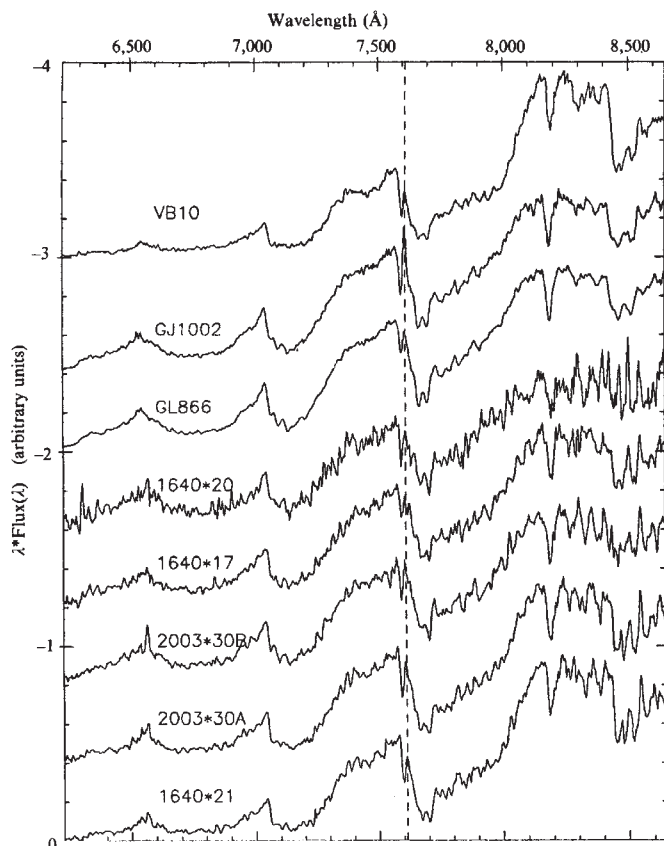


Fig. 1 Spectra of three very low luminosity M dwarfs with well determined absolute magnitudes (VB10, $M_v = +18.5$; GJ1002, $M_v = +15.5$; GL866, $M_v = +14.6$), and five representative field stars from the survey described in the text. Comparison of the luminosity-sensitive features in the spectra, particularly the spectral gradient from 8,000 to 8,150 Å, and the depth of the broad absorption feature near 8,450 Å, shows the field stars to have absolute magnitudes in the same range as the comparison stars. This allows the derivation of the distances, and therefore the space density, of these stars. Significant H α (6,563 Å) emission is seen in all these low-luminosity stars. The stars labelled 2,003 * 30A and 2003 * 30B are also of interest, as they are separated by only 10 arc s, and may form a binary system. The vertical dashed lines mark features which have been distorted by atmospheric A band absorption.

The local space density of luminous stars and stellar remnants is $\sim 0.05 M_\odot \text{ pc}^{-3}$, with a similar contribution from interstellar material⁴. This implies a local 'missing' mass of $0.05 M_\odot \text{ pc}^{-3}$ or about one low-luminosity M dwarf pc^{-3} . This value is two orders of magnitude above the observational limit derived here and comfortably in excess of the likely uncertainties; it implies that M dwarfs do not make up the local missing mass. The local space density of the global (flat rotation curve) missing mass is poorly determined, but may also be usefully constrained by these data. A recent estimate⁴, which is typical of published values, is $0.01 M_\odot \text{ pc}^{-3}$, equivalent to ~ 0.1 low-mass M dwarfs pc^{-3} . This value is a factor of 10 above our upper limit, a difference which is formally significant. There are, however, substantial uncertainties in the derivation of the local density of the large scale missing mass, so that this limit is less severe than that on the local missing mass.

Taken together, however, the results of this survey show that the large-scale missing mass is unlikely to be made up of low-mass luminous stars, while the solar neighbourhood missing mass cannot be made up of low-mass luminous stars.

G.G. is a visiting associate, Mount Wilson and Las Campanas Observatories, Carnegie Institute of Washington. Both G.G. and P.H. are visiting astronomers, Cerro Tololo Interamerican Observatory which is supported by the NSF under contract no. AST 78-27879.

Received 25 August; accepted 12 October 1983.

1. Faber, S. M. & Gallagher, J. S. *A. Rev. Astr. Astrophys.* **17**, 135-187 (1979).
2. Bahcall, J. N. *Astrophys. J.* (in the press).
3. Hill, G., Hilditch, R. W. & Barnes, J. V. *Mon. Not. R. astr. Soc.* **186**, 813-829 (1979).
4. Schmidt, M. *IAU Symp.* No. 106 (in the press).
5. Gilmore, G. & Reid, N. *Mon. Not. R. astr. Soc.* **202**, 1025-1047 (1983).
6. Reid, N. & Gilmore, G. *Mon. Not. R. astr. Soc.* **201**, 73-94 (1982).
7. Reid, N. & Gilmore, G. *Mon. Not. R. astr. Soc.* (in the press).
8. Gliese, W. & Jahreiss, H. *Astr. Astrophys. Suppl. Ser.* **38**, 423-448 (1979).
9. Greenstein, J. L., Neugebauer, G. & Becklin, E. L. *Astrophys. J.* **161**, 519-531 (1970).
10. Reid, N. & Gilmore, G. *Mon. Not. R. astr. Soc.* **196**, 15P-18P (1981).
11. Liebert, J. & Ferguson, D. H. *Mon. Not. R. astr. Soc.* **199**, 29P-30P (1982).

Sunspot index infers a small relict magnetic field in the Sun's core

C. P. Sonett

Department of Planetary Sciences and Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

Cowling¹ long ago suggested the presence of a relict field in the Sun's core and argued that the bulk electrical conductivity was sufficient (provided that the field was dipolar) to prevent decay over the lifetime of the Solar System. Evidence for such a field is obtained by inference from the asymmetry in time of the solar sunspot number taken over many cycles². As the Sun's dipole field reverses polarity approximately every 11 yr, if a core field of odd multipolarity (presumably dipolar) is present and in reasonable alignment with the oscillating dipole of the convective zone, it follows that in successive 11-yr periods, the two fields will tend first to be additive and then subtractive^{3,4}. Because the net dipole field is the 'seed' field from which the much larger toroidal field arises⁵, the latter will vary in magnitude in alternating half cycles and correspondingly mediate the level of sunspot activity. Whether a fossil or relict field is present can be explored either by a search for the presence of a 22-yr line quintet in the Fourier spectrum of the sunspot index², or alternatively the time sequence itself can be processed so as directly to bring out evidence for an offset. The latter involves a 'demodulation' of the sunspot sequence to emphasize the basic 22-yr Hale cycle. The latter is discussed here, but as the two are interrelated, the former is reviewed briefly first to provide the basis for the subsequent least squares (LS) analysis of the demodulated signal. As the solar cycle is unstable to drifts in frequency, the presence of frequency modulation can lead to a spurious relict field signal. A short summary of limits on frequency modulation (FM) contamination of the sunspot index spectrum is also given.

The lines in the spectrum of the Zurich sunspot index in the interval $100 < t < 8$ yr (where t is the period) match those of a model in which a 22-yr carrier 'wave' representing the Hale 22-yr cycle is amplitude modulated (AM) at the 90 yr (Gleissburg) period in the presence of a small but significant offset from zero mean of the 22 yr carrier, the entire function being squared. The 45-yr second harmonic of the Gleissburg period and line quintets about the 22- and 11-yr periods arise because the non-negative sunspot spectrum is a quadratic function of the modulated and offset carrier. The functional form of the model time sequence is given by²

$$f(t) = [1 + \alpha \sin(\omega_m t)][\Delta + \cos(\omega_c t)]^2 + n(\rho)^2 \quad (1)$$

where $\omega_m = 1/90 \text{ yr}^{-1}$ is the Gleissburg modulation frequency, $\alpha = 0.25$ the amplitude modulation index, $\omega_c = 1/22 \text{ yr}^{-1}$ the Hale 22 yr 'carrier' frequency, $\Delta = 0.05$ the offset in the Hale carrier, and $n(\rho)$ a gaussian noise term of 5% squared amplitude.

Demodulation of the sunspot time sequence is carried out by an 'unfolding' operation which may be thought of as the inverse of rectification of a modulated sine wave by squaring. Extraction of the square root of the sequence is followed by reversing the sign of alternating half cycles. The conceptual justification is

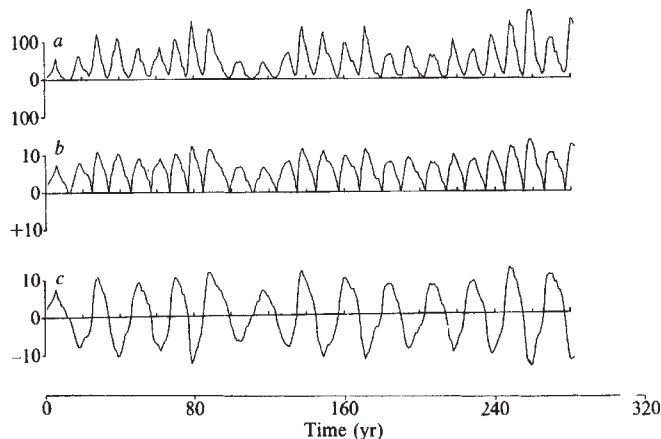


Fig. 1 *a*, The 'raw' sunspot time sequence for 1700–1982 (adapted from ref. 20, except for the past 20 yr from Eddy, personal communication); *b*, the square root (unsquared) sunspot index time sequence; and *c*, the latter with alternating polarity.

that the observed sunspot sequence is non-negative, but the underlying Hale cycle alternates in algebraic sign. As the sunspot time sequence does not generally attain zero value at sunspot lows, the unfolding is imperfect and the function must be slightly modified, forcing the minima to zero. This is consistent with the observation that magnetic half cycles on the Sun appear to overlap, that is a new half cycle is born before complete disappearance of the previous half cycle of opposite polarity. Figure 1*a* shows the raw sunspot sequence compared with the 'unsquared' index (Fig. 1*b*), and the polarity alternated version in Fig. 1*c*.

The multivariate LS fitting of the Hale cycle, assumed to be the dominant physical period of the convective zone dynamo, is carried out by fitting the function

$$g(t) = k_0(t) + k_c(t) \cos \omega_c t + k_s(t) \sin \omega_c t \quad (2)$$

where $k_0(t)$, $k_c(t)$, and $k_s(t)$ are free parameters to be fitted under the LS constraint. The LS calculation is made over 45 data points representing exactly two cycles of ω_c ; it is then stepped by 1 yr and repeated. The time dependence of the k values denotes their dependence on the particular LS calculation. Although the value of ω_c is obtained from a maximum entropy (MEM)^{6,7} estimate of this dominant period, contamination of the estimates of k_0 , k_1 , and k_2 is possible, as the sequence may retain other spectral components. To reduce this possibility for periods of the order of 11 yr or shorter, we first apply a centre weighted filter (with unity centre amplitude linearly tapered to zero at ± 22 yr about the centre value) to the sequence before the LS computation. This filter is run over the entire time sequence of 282 yr in 1-yr increments (with the filter centre weight started and ended 11 yr from the ends). The sequence is additionally truncated at each end by 11 yr to eliminate the asymmetrically weighted sequence ends. The orthogonality of the terms in equation (2) is demonstrated by the covariances $r_{o,s} = 0.00$, $r_{o,c} = 0.00$, and $r_{c,s} = 0.02$ where the subscripts are related to the coefficients of equation (2). These are evaluated for each LS calculation. Their time invariance and small values indicate that orthogonality of the fitting functions of equation (2) is maintained through the series of progressive LS calculations.

The computation yields a time varying zero offset, phase, and amplitude. The first two parameters are shown in Figs 2 and 3 with associated error limits. As can be expected because the sunspot time sequence is not stationary in time, these parameters display limited stability over the record of 260 yr length, with a key exception in the 1780–1800 period. At that time a large-scale perturbation took place which affected all three parameters. Dicke⁸ has previously called attention to this unusual period of time which may, however, be associated with errors in the Wolf record⁹ (also J. A. Eddy and D. V. Hoyt, personal

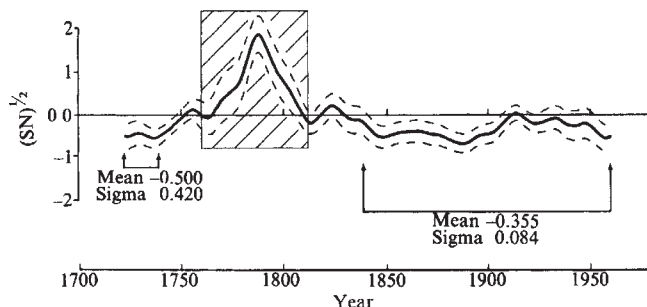


Fig. 2 The multivariate least squares residue for offset (k_0) of the carrier derived from Fig. 1*c*. The solid line is the weighted sliding mean over 45 yr, and the dashed lines are the corresponding variances. The data shown here has been truncated by 11 yr at each end. The unconformity in the 1780–1800 period is shown with superimposed cross-hatching, and is shown primarily for completeness of the record.

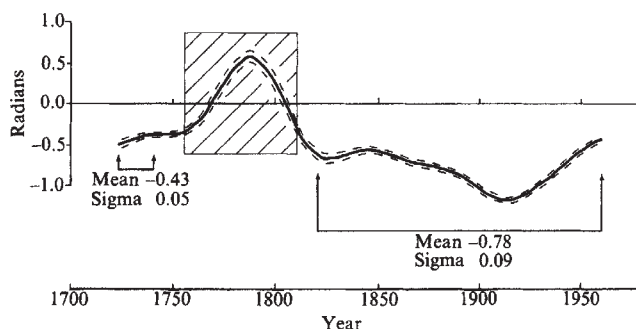


Fig. 3 The companion carrier phase from the LS residues. This figure corresponds to the phase derived from the coefficients, k_c and k_s of the LS fit. Other aspects are as for Fig. 2.

communication). This part of the record is indicated by a superimposed cross-hatching, and is retained here only for completeness.

The anomalous period at the end of the eighteenth century, if included in the analysis, would seriously bias estimates of the offset towards a smaller value, although even with its (we believe) improper inclusion in the overall statistics the average offset would still retain a negative value. Accordingly, the record has been divided into the short section just before this period and the sub-sequence of 140 yr length following the anomaly, and the anomalous period is discarded. The two non-contiguous sub-sequences each give a significant mean negative offset. Attention is centred on the latter because of its greater length. For this period the mean and standard deviation are -0.36 and 0.08 in units of sunspot index^{1/2}.

Inspection of Fig. 2 discloses variations about a mean offset. The variations are hardly surprising because the time sequence was low pass filtered and as we noted earlier the sequence contains significant power in several peaks of period longer than 22 yr. An MEM estimate of the residual spectral content shows that k_0 contains a 200 yr period. Consequently we have carried out a further LS computation on the offset, using a repeat of the constraints given by equation (2), but with the period equal to 200 yr. The calculation is done with the unconformity removed. This leaves a 50-yr gap in the sequence to which the 200 yr calculation is relatively insensitive. The outcome of this second moving LS computation yields $k_0 = -3.19$, $k_c = -1.58$, and $k_s = -0.097$, with the standard deviation of the offset k_0 equal to 0.14 . The final outcome is composed about two-thirds of a steady component and one-third time varying part with $t = 200$ yr.

The phase (Fig. 3) shows a mean value of about 0.5 rad (not especially important as the reference is arbitrary), but deviations

from the mean are as great as 1 rad (for the 'unconformity' of AD 1780–1800) and less at other times. The phase time rate has a maximum of about 0.03 rad yr^{-1} or 0.0047 yr^{-1} with peak frequency modulation (FM) at a period of about 90 yr. This corresponds to an FM modulation index of 0.1 and a deviation ratio of 0.40.

The major source of FM is centred about the unconformity ~1780–1800, where the carrier period varies from 24.9 yr downward to 20.1 yr, a variation of about 10% somewhat less than the estimate of Eddy¹⁰. Frequency components higher than about 0.05 yr^{-1} are removed by the low pass filter we use, so that may account for some of the disparity.

The generalization of equation (1) to include frequency modulation is given by:

$$g(t) = [1 + \alpha \cos \omega_m t][\Delta + \cos(\omega_c t + \beta \cos \omega_t t)]^2 + n^2(\rho) \quad (3)$$

where the parameters are those defined for equation (1) with the addition of the FM angular frequency ω_t and the deviation ratio β . Expansion of equation (3) yields a complicated spectrum whose terms can be grouped by dependence on Δ , that is

$$g(t) = A(t) + \Delta B(t) + \Delta^2 C(t) \quad (4)$$

where $A(t)$ contains terms centred on $2\omega_c$, $B(t)$ terms centred on ω_c , and $C(t)$ terms associated only with ω_m . The linear dependence of $B(t)$ on Δ is given by

$$B(t) = 2\alpha \Delta \cos[\beta \cos(\omega_t t) - \sin[\beta \cos(\omega_t t)] \times [\cos(\omega_c + \omega_t)t + \cos(\omega_c - \omega_t)t] \quad (5)$$

The lines associated with $A(t)$ are centred on the second harmonic of the Hale cycle; they are expected to contribute only weakly in the neighbourhood of the Hale cycle itself. Those lines associated with $B(t)$ are in a band centred on ω_c ; because $B(t)$ is linearly dependent on Δ , it provides a measure of the offset field. Those lines associated with $C(t)$ are quadratically dependent on Δ , a small number, and therefore relatively unimportant. Using the phase variation determined from the first LS analysis of the complete time series (including the unconformity of 1780–1800) $\alpha = 0.25$, the FM deviation ratio, $\beta = 0.4$, and the mean carrier period is 22.15 yr. For an amplitude modulation period of 90 yr, the FM side band power and periods can, using the term $B(t)$ of equation (4), be shown not to be important.

The presence of a fossil field suggests a residue from the initial contraction of the pre-main-sequence Sun, drawn together from the field of the pre-solar nebula from which the Solar System formed. Computation of the intensity of that field is made difficult by the uncertain source—a fragment of the background field of the local arm of the galaxy, the field of a pre-solar local system dynamo, a pre-main phase solar dynamo, or some combination of these¹¹. The extreme magnetic pressure would be expected to oppose the solar condensation, although turbulent dissipation might decrease the field intensity and reduce the opposition to condensation.

An estimate of the putative relict core field is based on the peak poloidal field of the Sun. Values range upwards from 1 G for the polar field to as high as 15 G from computation of the field based on magnetic potential models using optical coronagraph data. Making the very simple but plausible association of the peak sunspot number with that in the poloidal field assumed to lie in the range of 1–15 G, the 4% offset arrived at from the series least squares analysis suggest a relict field of 0.04–0.6 G. This is orders smaller than for the condensed primordial field, assuming no turbulent dissipation. This the present relict field would be only a small residue of the original field, consistent with a strong early loss of field.

Levy and Boyer³ and Boyer and Levy⁴ discuss the role that a fossil dipolar core field would have on the solar cycle, yielding "a systematic variation in the vigor of solar activity, or some other property of the sunspot cycle". Their calculation suggests that a fossil field of order 0.1 G could so modify magnetic solar activity that the alternating half cycles might vary in strength

by a factor of 3. Note that the calculation bearing on this is one-dimensional and only illustrative. A more realistic three-dimensional calculation is now under way (E. H. Levy, personal communication).

The sweeping together of magnetic field by the condensing Sun seems likely to collect a relatively well ordered set of field lines, but we cannot be certain of this. The addition of turbulence suggests that the final configuration of field might contain orders higher than dipole. The decay of the transverse magnetic (TE) mode for a sphere is given by^{12,13}

$$t_n = \mu \sigma r^2 / (\pi n)^2 \quad (6)$$

where t_n is the n th order decay time constant, σ is the bulk electrical conductivity, μ the permeability taken as that of free space, r the radius, and n the order. Thus higher orders decay with time constant varying as $1/n^2$. Using the radius of the Sun for r , the first-order decay time is $5 \times 10^5 \text{ yr}$, while higher orders decay correspondingly more rapidly. Thus the relict core field geometry should be primarily dipolar; its orientation should depend on the original orientation of the magnetic field of the pre-solar nebula. The field would have been aligned with the solar spin axis if the latter has not changed and if the nebular field contained a dynamo component. D. Boyer (personal communication) has pointed out that higher-mode steady fields might be present, but this is uncertain. If so it would indicate that though the decay is faster than for the dipole mode, a very high initially captured field might still yield a present epoch higher mode residue.

Arguments why the relict field cannot be very large¹¹, contrast with a simple model of condensation alone of an interplanetary field of 1 G where one might expect a final core field amplified to perhaps 10^5 G . Suggestions of a large field in the solar core have been recently made based on observation of bulk oscillations of the Sun¹⁴. Whether such fields, generally associated with a core spinning more rapidly than the convective zone, could lead to the alternation in sunspot amplitude has not been investigated. But it seems difficult to reconcile the typical oscillatory periods of such a core with the very long periods arrived at here, although evidence for a radial oscillation of long period has recently been reported^{15–17}. Gilliland's value is $76 \pm 8 \text{ yr}$ (ref. 15).

Controversy still surrounds the large field hypothesis¹⁸. Dicke discusses the possibility of a core field being aligned normal to the solar spin axis¹⁹, resulting in containment because of a shell of differential rotation between the core and convective zone. As the normal component of the core field should not be affected by shear rotation and will appear in the zone of convection, it will be in contact with the region in which the solar dynamo is generated. Thus such a field may not be observationally separable from an axial field.

I thank J. A. Eddy for making available the 20-yr sunspot record extending to 1982. L. D. Smith carried out much of the analysis of the time series. I am also indebted to E. H. Levy and to D. Boyer for discussions of their work. B. A. Hobbs noted the residual cyclicity in the offset. This work was supported partially by NASA grant NSG 7024 and by NSF grant ATM-8308378.

Received 10 January; accepted 20 September 1983.

1. Cowling, T. G. *Mon. Not. R. astr. Soc.* **105**, 167–174 (1945).
2. Sonett, C. P. *Geophys. Res. Lett.* **9**, 1313–1316 (1982) in *Weather and Climate Responses to Solar Variations*, 607 (ed. McCormac, B. M.) (Colorado Associated University Press, 1983).
3. Levy, E. H. & Boyer, D. *Astrophys. J. Lett.* **254**, L19 (1982).
4. Boyer, D. & Levy, E. H. *Astrophys. J.* (in the press).
5. Parker, E. N. *Astrophys. J.* **156**, (1970).
6. Box, G. E. P. & Jenkins, G. M. *Time Series Analysis* 2nd edn (Holden-Day, San Francisco, 1976).
7. Ulrich, T. J. & Bishop, T. N. *J. geophys. Res.* **13**, 183 (1974).
8. Dicke, R. H. *Nature* **276**, 676 (1978).
9. Sonett, C. P. *J. geophys. Res.* **88**, 3225 (1983).
10. Eddy, J. A. in *The Ancient Sun* (eds Pepin, R. O., Eddy, J. A. & Merrill, R. B.) (Pergamon, Oxford, 1980).
11. Levy, E. H. & Sonett, C. P. in *Protostars and Planets*, 516 (ed. Gehrels, T.) (University of Arizona Press, 1978).
12. Smythe, W. R. *Static and Dynamic Electricity* 3rd edn (McGraw-Hill, New York, 1968).

13. Sonett, C. P., Mihalov, J. D. & Ness, N. F. *J. geophys. Res.* **76**, 5172 (1971).
14. Gough, D. O. *Nature* **298**, 350–354 (1982); *Nature* **298**, 334–339 (1982).
15. Gilliland, R. L. *Astrophys. J.* **248**, 1144 (1981).
16. Parkinson, J. H. *Nature* **304**, 518 (1983).
17. Sofia, S., Dunham, D. W., Dunham, J. B. & Fiala, A. D. *Nature* **304**, 522 (1983).
18. Isaak, G. R. *Nature* **296**, 130–131 (1982).
19. Dicke, R. H. *A Rev. Astr. Astrophys.* **8**, 297–328 (1970).
20. Vitinskii, Yu. I. Preprint, Academy of Sciences, USSR, Leningrad (1962) (translated by Israel Program for Scientific Translation, Jerusalem (1965) [NASA TTF-289]).

Hydrogen chemical potentials and diffusion coefficients in hydrogen diffusion membranes

F. A. Lewis, J. P. Magennis, S. G. McKee
& P. J. M. Ssebuwufu*

Chemistry Department, Queen's University, Belfast BT9 5AG, UK

Palladium and palladium alloys are quite widely used as membranes for hydrogen permeation, because of their resistance to embrittlement and to irreversible deformations during cycles of absorption and desorption of hydrogen^{1–3}. The isothermal relationships between hydrogen diffusion coefficient, D_H , and hydrogen content, n (ratio of hydrogen/metal atoms) has both technological and academic interest. The most detailed form of relationships^{2,3} between D_H and n have been those reported for deuterium in $\text{Pd}_{75}\text{Ag}_{25}$ by Hickman⁴ within the temperature range 300–500 °C, and for protium in $\text{Pd}_7\text{Ag}_{23}$ within the range 30–75 °C by Küssner⁵. For increasing values of n up to $n \sim 0.2$, Küssner⁵ reported significant corresponding decreases of D_H (by a factor of ~ 8 from $n = 0$ up to $n \sim 0.16$ at 30 °C). This is the reverse of trends in the values of D_H with increasing n suggested by measurements of anelastic effects^{3,6,7} and also by the permeation rate measurements of Hickman⁴ at 300 and 400 °C. Here we report measurements of diffusion through $\text{Pd}_{81}\text{Pt}_{19}$ alloy, the results of which are consistent with Gorskii's hypothesis relating long-range anelastic diffusion of lattice interstitials to elastic strain gradients.

In Küssner's studies⁵, values of D_H , were calculated from 'breakthrough times' estimates^{2,8} from plots of the time dependence of electrode potentials measured at the diffusion (hydrogen exit) side of a $\text{Pd}_7\text{Ag}_{23}$ bielectrode, Küssner also reported that after each increase of electrode potential imposed on the hydrogen input side of the bielectrode, consequent changes of the electrode potential at the diffusion (exit) side, exhibited an initial period of alteration indicative of decreases of the surface hydrogen chemical potential μ_H , before the onset of the period of increasing μ_H which could seem to have been expected to correspond with increases of n .

Küssner associated the initial changes of electrode potential with changes of μ_H arising from compressional deformation effects, allied to some measure of irreversible damage to the membranes. A possibility was inferred in a subsequent generally analogous study with Pd–Au alloys⁹ that Küssner's observations of apparent initial decreases of μ_H , might have been attributable to artefacts associated with the electrical circuitry.

Since the studies of Küssner⁵ there have been experimental demonstrations^{2,3,6,10–16} of the postulate by Gorskii¹⁷ of the long-range anelastic diffusion of unlike lattice entities, such as hydrogen interstitials, under the influence of elastic strain gradients produced by mechanical stresses^{6,10–14}, or in the virtual absence of phase transitions, by the relative differences of lattice expansions involved in hydrogen concentration gradients^{15,16}. Also, the general possibility has been pointed out¹⁸, that Gorskii effect phenomena could influence values of diffusion coefficients determined by breakthrough techniques.

At 25 °C the form of the relationships between hydrogen chemical potential and hydrogen content (μ – n relationships) of the $\text{Pd}_{81}\text{Pt}_{19}\text{H}_n$ system are basically analogous to those of the

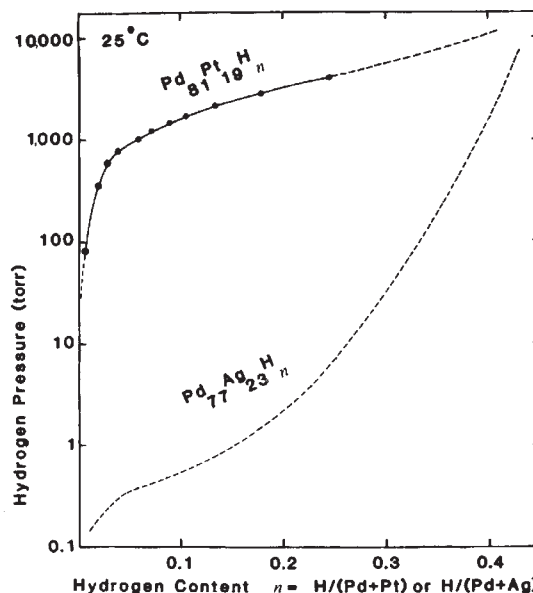


Fig. 1 Relationships between equilibrium hydrogen pressures and hydrogen contents for $\text{Pd}_{81}\text{Pt}_{19}$ and $\text{Pd}_{77}\text{Ag}_{23}$ alloys at 25 °C.

$\text{Pd}_{77}\text{Ag}_{23}\text{H}_n$ system. However, in the $\text{Pd}_{81}\text{Pt}_{19}\text{H}_n$ system^{19,20}, the 'inflected' region of the relationship, where n increases most substantially with increasing μ_H , corresponds to substantially higher equilibrium pressures than in the case of $\text{Pd}_{77}\text{Ag}_{23}\text{H}_n$ (or $\text{Pd}_{75}\text{Ag}_{25}\text{H}_n$). This is indicated in Fig. 1 where the form of relationship for $\text{Pd}_{77}\text{Ag}_{23}\text{H}_n$ has been estimated from data^{5,21,22} at adjacent temperatures and closely similar Pd/Ag compositions.

The tubes of $\text{Pd}_{81}\text{Pt}_{19}$ alloy had a wall thickness of 0.2 mm. They were used in two types of experiments: (1) the insides of the tubes were filled with an aqueous solution of sulphuric acid which was generally pre-saturated with hydrogen gas at atmospheric pressure followed by a period of quiescence before experiments. A platinized platinum reference electrode was immersed in the solution in the middle of the tube. In these experiments, the electrolyte solution adjacent to the outside wall also was similarly pre-saturated with hydrogen and the inner surface of the tube was always coated with palladium black. (2) The tubes were connected to a gas line, which permitted evacuation, measurements of hydrogen pressures within the tube by Baratron (type 222A) capacitance manometers, and the introduction of other gases.

In both types of experiments, the outer walls of the tubes were coated by electrodeposition of palladium black, and could act as either the cathode or anode of an electrolytic cell at current densities generally of $\sim 15 \text{ mA cm}^{-2}$. In either case it could be expected^{3,20,23,24}, that within 1 min of commencement of electrolysis surface hydrogen concentrations corresponded to $n \sim 0.4$ during cathodization (see Fig. 1) and $n \sim 0$ during anodization.

In the case of the type 1 experiments, tube walls had an initial hydrogen content of $n \sim 0.04$ (refs 20, 24) (see Fig. 1). Typical forms of subsequent changes of the electrode potential, E , of the inner wall, with respect to the internal Pt/H_2 reference, after cathodizations and anodizations of the outer wall respectively, are illustrated in Fig. 2.

In the case of experiments of type 2, the tubes were first evacuated and hydrogen gas generated within them by cathodization of the outer wall. Following interruptions of electrolysis, hydrogen pressures within the tubes settled to almost constant values governed kinetically (for inner surfaces coated with palladium black) by molecular diffusion of hydrogen within the interfacial solution layers adjacent to the outer wall^{23,24}. Typical forms of changes of the hydrogen pressure, p , within the tubes subsequent to recommencement of either cathodizations or anodizations of the outer wall are also illustrated in Fig. 2. In

* Present address: Chemistry Department, Makerere University, Kampala, Uganda.

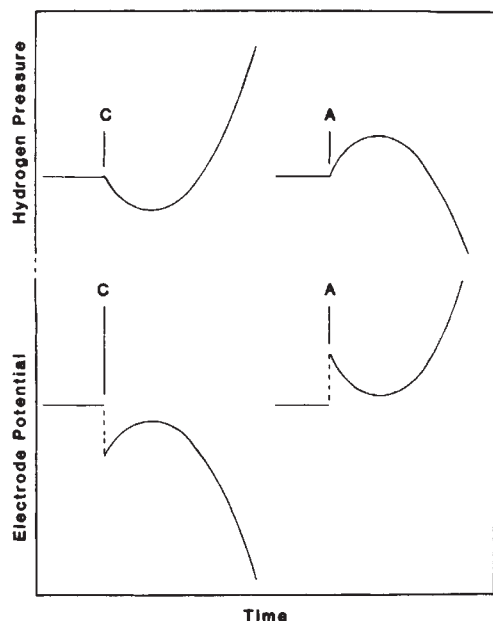


Fig. 2 Diagrammatic representations of patterns of changes with time, at 25 °C, for Pd₈₁Pt₁₉ tubes with palladium-black coated inner walls, of: electrode potentials at the inner walls (type 1 experiments) and of changes of pressures of hydrogen gas contained within them (type 2 experiments) following cathodic hydrogen discharges, C, and anodic hydrogen removals, A, at their outer walls—after previous initial establishments of almost steady-state conditions with regard to the hydrogen contents of the tube walls.

the type 1 experiments initial abrupt changes of values of E at the inner wall occurred in <10 ms and seemed clearly to originate within the electrical circuitry. Subsequent increases of E ($+\Delta E$) on cathodization and decreases on anodization ($-\Delta E$) required 10–15 min to attain maximum and minimum values respectively. Assuming ΔE to be indicative of changes of μ_{H} , and equatable to corresponding equilibrium hydrogen pressures p at temperature $T = 298$ K by

$$2FE = RT \ln (p/P) \quad (1)$$

where F and R are the Faraday and gas constants respectively and with P taken as 1 bar (10^5 Pa), typical corresponding decreases of p following cathodization and increases of p following anodization were calculated to be close to 50 and 20 torr respectively. In experiments of type 2, the initial decreases of p ($-\Delta p^*$) following cathodization and increases ($+\Delta p^*$) following anodization (of the outer wall in both cases) occurred over time intervals of similar magnitude to the analogous forms of changes of ΔE . Features of these pressure change effects were as follows:

First, effects were almost imperceptible if the inner surface of the tube was inactive for equilibration with hydrogen molecules, such as after treatment with iodine³. Second, even with inner surfaces activated by palladium black, as below, no comparable effects were observed if argon at various pressures was substituted in the tubes although the tube walls still retained a hydrogen content closely corresponding to the various previously developed equivalent hydrogen pressures. Third, magnitudes of effects illustrated in the upper part of Fig. 2, markedly increased when inner surfaces of the tubes were also coated with palladium black; to improve correspondences between the hydrogen chemical potential at the internal surface and the internal hydrogen pressures. In such experiments maximum extents, Δp^*_{max} , of initial decreases of pressure following recommencements of cathodization markedly increased with increasing values of the initial hydrogen pressures in the tubes, and thus the initial hydrogen contents, n , of the tube walls. There were also accompanying lengthenings of time intervals before reincreases of hydrogen pressure occurred after the initial peri-

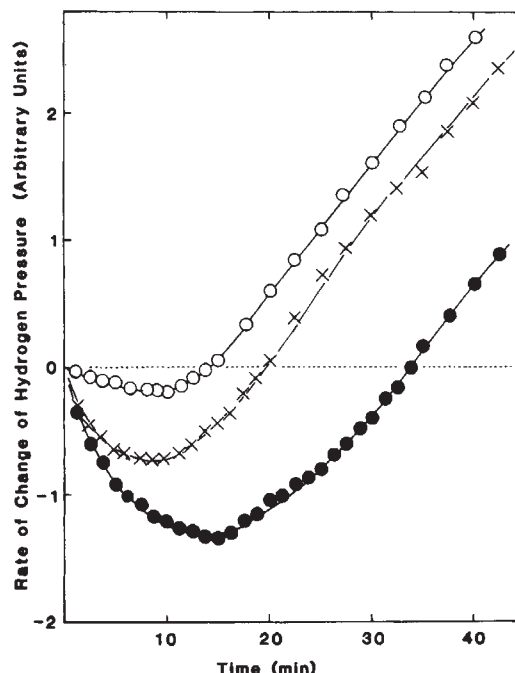


Fig. 3 Comparison of rates of change of hydrogen pressures (negative values correspond to pressure decreases) at 25 °C, within Pd₈₁Pt₁₉ tubes with palladium-black coated inner walls, following cathodic discharge of hydrogen at ~ 15 mA cm⁻² at their outer walls after establishments of initial virtually time-independent pressures of hydrogen within the tubes of: ○, 189; ×, 569; ●, 748 torr respectively.

ods of decrease. For example, in Fig. 3, times of onset of positive rates of pressure changes corresponded to ~ 14 , ~ 20 and ~ 34 min respectively, following previous establishments (before recommencements of cathodization) of almost constant hydrogen pressures of ~ 190 , ~ 570 and ~ 750 torr respectively, which from Fig. 1 correspond to initial hydrogen contents of the tube wall of $n \sim 0.016$, $n \sim 0.027$ and $n \sim 0.040$ respectively. The maximum reductions of initial pressures, $-\Delta p^*_{\text{max}}$, corresponding to the minima in Fig. 3, were ~ 5 , ~ 10 and ~ 25 torr respectively.

The degree of correspondence between the forms and magnitude of pressure-time relationships derived from values of E , and those of the directly measured gaseous pressure against time relationships has provided self-consistent evidence that all effects represented in Fig. 2, apart from the initial abrupt changes of electrode potential, are indicative of the course of changes of μ_{H} at the inner surface of the tubes. This conclusion is essentially in keeping with Küssner's interpretation⁵ of his measurements of electrode potential, but the measurements of pressure changes provide correlating evidence that at a sufficiently catalytically active surface, the changes of μ_{H} lead to developments of corresponding equilibrium hydrogen pressures.

These pressure changes have, however, involved substantial transfer of hydrogen between the inner wall of the tube and the gas within it, which initially occur in the opposite direction to the electrolytic inputs or removal of hydrogen through the outer surface. Thus taking the case of increases of n of the outer wall, during cathodization, there is an initial consequent transfer of hydrogen from the gas within the tube into the inner wall. Although this might simply be interpreted as producing an increase of the hydrogen content of the inner wall this does not seem compatible with the corresponding decreases of hydrogen chemical potentials, μ_{H} , shown by the electrode potential. These decreases of μ_{H} would, however, logically provide an impetus for absorption of hydrogen gas, to re-establish thermodynamic equilibrium.

Such initial reductions of the inner wall surface hydrogen chemical potential and accompanying entry of hydrogen gas

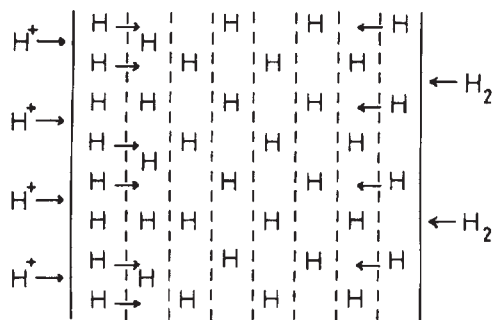


Fig. 4 Representation of the cross-section of a metal membrane initially containing hydrogen, during early stages of hydrogen discharge at the left-hand face, and accompanying relative expansions of the adjacent planes of metal atoms resulting from increases of hydrogen content and Gorskii effect strain-induced hydrogen diffusion from planes of metal atoms adjacent to the right-hand face, accompanied by compensating entry of hydrogen gas through this catalytically active right-hand surface.

from within the tube can be accounted for in terms of the Gorskii effect. Thus initial expansions of the outer wall relative to the inner wall produced^{1-3,25} by entry of hydrogen during cathodization can cause strain gradients^{15,16} which will induce movement of hydrogen from regions of the metal lattice adjacent to the inner wall towards the outer wall. This is illustrated in Fig. 4. For approximately the same initial hydrogen concentration differences between outer and inner walls, and thus similar strain gradients, the magnitude of hydrogen movements due to the Gorskii effect, can be expected to be proportional to the much smaller initial concentrations of hydrogen within the inner wall of the tube relative to the initial concentration in the outer wall as has been observed in the present studies.

The converse effects, produced following removals of hydrogen from outer walls of the tubes by anodization, of initial increases of the hydrogen chemical potential of the inner wall as indicated by initial increases of hydrogen gas pressures and by the direction of changes of electrode potential—can be analogously accounted for by rather small initial strain gradients in the reverse direction with hydrogen concentration gradients corresponding to values of n in the present experiments of $0 \approx n \leq 0.04$ as compared with $0.4 \approx n \leq 0.04$ in the cases of cathodization of the outer wall.

If Gorskii effect induced hydrogen diffusion within the tube wall is responsible for the main effects illustrated in Fig. 2, rather important points of principle require reconsideration both directly in regard to the Gorskii effect and to the determinations of relationships between D_H and n by breakthrough techniques.

For example, it follows from results of type 2 experiments, that changes of the hydrogen content of the outer wall produce overall movements of the hydrogen from within regions of the tube wall adjacent to the inner surface, in an opposite direction to the overall gradient of hydrogen chemical potential—when this is considered with respect only to values of μ_H at the inner and outer surfaces.

From this it would seem reasonable to infer in other examples of the Gorskii effect in metal-hydrogen systems, that the overall transfer of hydrogen can be expected to be occurring in an opposite direction to gradients of hydrogen chemical potential. It also follows that, in regions of the metal lattice which are relatively compacted by imposed stresses producing Gorskii effect phenomena, the final hydrogen chemical potential will have undergone a reduction by the operation of the effect. This general conclusion is opposite to assumptions that hydrogen migrates from compacted regions of the metal lattice, under the impetus of increased hydrogen chemical potentials and does not seem to be in keeping with suggestions that the hydrogen in metal-hydrogen systems can be better considered as a lattice gas^{12,26-28} rather than treating these systems as an important group of a wider class exhibiting nonstoichiometry²⁹.

With reference to the general diffusion picture, an initial induced diffusion of quantities of hydrogen in the opposite direction to that of the final steady-state flux constitutes an example of a 'non-Fickian' diffusion situation^{30,31}, and can be expected to affect the interpretation of 'breakthrough times' and derived values of hydrogen diffusion coefficients.

The present results have shown that extents of effects generalized in Fig. 2 are increased both by increases of the initial mean hydrogen content of the membrane and by increases of the initially imposed differences of hydrogen content between the two surfaces.

In the experiments of Küssner⁵, no corrections with regard to calculations of D_H were applied in regard to either of these factors. Moreover, in regard to the latter factor of initial concentration gradient (and thus strain gradient) differences, Küssner would appear⁵ to have applied very similar incremental increases of the electrode potential of the 'hydrogen-input' wall at each successive value of n . As may be appreciated from Fig. 1, these gradients would have been expected to increase markedly over the initial regions of increasing μ_H with increasing n , and would be particularly large over the 'inflected' region of the μ_H - n relationship where μ_H increases very little in proportion to increases of n .

Because of either of these two factors, 'breakthrough times', derived on assumption of simple Fickian diffusion conditions, and defined^{2,4,5,8} as times corresponding to intersections with the time axis of plots (over regions where hydrogen is clearly evolving at the diffusion side of membranes) of E or p , derived functions of these, or other related experimental parameters such as in Fig. 3, can be substantially too high and thus derived values of D_H correspondingly too low. It now therefore seems a strong possibility that the lack of appreciation of influences of Gorskii effect phenomena in Küssner's analysis of his measurements⁵, could provide an explanation of indications²⁻⁸ that D_H decreases significantly with increasing n up to ~ 0.16 (30–75 °C) in Pd₇₇Ag₂₃ alloys, rather than the continuous increases of D_H with increasing n which have been suggested by results of studies of anelastic effects^{3,6,7} and have also been generally indicated by permeation rate studies⁴ at higher temperatures.

We acknowledge Johnson, Matthey & Co. for provision of the Pd₈₁Pt₁₉ alloy tubes, SRC for partial financial support, and also J. H. Kirkpatrick and S. E. J. Bell for technical assistance. S. G. McKee also acknowledges awards of Musgrave and Larmor Studentships by Queen's University.

Received 13 June; accepted 12 October 1983.

- Lewis, F. A. *The Palladium-Hydrogen System* (Academic, London and New York, 1967).
- Wicke, E., Brodowsky, H. & Züchner, H. *Topics appl. Phys.* **29**, 73 (1978).
- Lewis, F. A. *Platinum Metals Rev.* **26**, 20, 70, 121 (1982).
- Hickman, R. G. *J. Less-Common Metals* **19**, 369 (1969).
- Küssner, A. *Z. Naturforsch.* **21a**, 515 (1966).
- Mazzolai, F. M., Nuovo, M. & Lewis, F. A. *Scripta metall.* **9**, 617 (1975).
- Nuovo, M., Mazzolai, F. M. & Lewis, F. A. *J. Less-Common Metals* **49**, 37 (1976).
- Boes, N. & Züchner, H. *J. Less-Common Metals* **49**, 233 (1976).
- Maestes, S. & Flanagan, T. B. *J. phys. Chem.* **77**, 850 (1973).
- Schaumann, G., Völkl, J. & Alefeld, G. *Phys. Rev. Lett.* **21**, 891 (1968).
- Cantelli, R., Mazzolai, F. M. & Nuovo, M. *Phys. Status Solidi* **34**, 397 (1969).
- de Ribapierre, Y. & Manchester, F. D. *J. Phys. C*, 2126 (1974).
- Völkl, J. & Alefeld, G. *Z. phys. Chem. neue Folge* **114**, 129 (1979).
- Mazzolai, F. M. & Züchner, H. *Z. phys. Chem. neue Folge* **124**, 59 (1981).
- Cermák, J. & Kufadakis, A. *J. Less-Common Metals* **49**, 309 (1976).
- Kufadakis, A., Cermák, J. & Lewis, F. A. *Surf. Technol.* **16**, 57 (1982).
- Gorskii, V. S. *Phys. Z. SowjU.* **8**, 457 (1935).
- Wipf, H. *J. Less-Common Metals* **49**, 291 (1976).
- Carson, A. W., Flanagan, T. B. & Lewis, F. A. *Nature* **183**, 39, 510 (1959).
- Baranowski, B., Lewis, F. A., McFall, W. D., Filipek, S. & Witherspoon, T. C. *Proc. R. Soc. A* **386**, 309 (1983).
- Brodowsky, H. & Poeschel, E. *Z. phys. Chem. neue Folge* **44**, 143 (1965).
- Carson, A. W. & Lewis, F. A. *Trans. Faraday Soc.* **63**, 1453 (1967).
- Lewis, F. A., Johnston, R. C., Witherspoon, M. C. & Obermann, A. *Surf. Technol.* **18**, 147 (1983).
- Lewis, F. A., Hull, M. N., Johnston, R. C. & Witherspoon, M. C. *Surf. Technol.* **18**, 167 (1983).
- Baranowski, B., Majchrzak, S. & Flanagan, T. B. *J. Phys.* **F1**, 258 (1971).
- Alefeld, G. *Phys. Status Solidi* **32**, 67 (1969).
- Manchester, F. D. *J. Less-Common Metals* **49**, 1 (1976).
- Bond, R. A. & Ross, D. K. *J. Phys.* **F12**, 597 (1982).
- see e.g., Burch, R. & Lewis, F. A. *A. Rep. chem. Soc.* **67**, 231 (1970).
- Petropoulos, J. H. & Roussis, P. P. *J. chem. Phys.* **47**, 1491, 1496 (1967); **48**, 4629 (1968); **50**, 3951 (1969).
- Cermák, J. *Rozn. Chem.* **50**, 1741 (1975).

Reaction of NO₂ with NaCl and atmospheric implications of NOCl formation

B. J. Finlayson-Pitts

Department of Chemistry, California State University, Fullerton, California 92634, USA

Oxidation of nitrogen dioxide in ambient air is of global interest as a source of O₃ in photochemical air pollution¹ and of HNO₃ in acid rain² and fogs³. Gas phase oxidation occurs by reaction with the photochemically produced hydroxyl free radical (OH) during the daylight hours². A second potential mechanism involves reactions of NO₃ (ref. 4). IR and mass spectrometric (MS) data are now reported for a third route which occurs both day and night, the reaction of NO₂ with sea salt (NaCl):



This is the first evidence that reaction (1), suggested previously^{5,6} proceeds at p.p.m. NO₂ concentrations. NOCl formation may be significant in coastal urban areas because it photolyses rapidly to Cl atoms, and hydrolyses to HONO, a major OH source; both Cl and OH initiate atmospheric photochemical oxidations.

An observed deficiency of chloride ions in coastal aerosols in some locations has been attributed to a reaction of HNO₃ or NO₂ with NaCl (refs 7–12). Gaseous HCl has been assumed to be the product based on studies^{11,12} showing that 0.1–100 p.p.m. NO₂ reacted rapidly with moist NaCl aerosols to produce NaNO₃ and gaseous HCl. More recently, nitrosyl chloride (NOCl) was identified⁶ from the reaction of large salt particles with torr pressures of NO₂, and NaNO₃ was observed in separate studies¹³, suggesting reaction (1). However, the reactive species may have been N₂O₄ which is present at substantial concentrations in equilibrium¹⁴ with NO₂ at torr pressures; in this case, atmospheric formation of NOCl would not be significant since negligible concentrations of N₂O₄ exist at ambient parts per 10⁶ or 10⁹ (p.p.m. or p.p.b.) NO₂ concentrations. N₂O₄ has in fact been shown¹⁵ to react with NaCl photochemically at low temperatures. Additionally, no stoichiometric data have been reported to support reaction (1).

We report here evidence for reaction (1) at room temperature and NO₂ concentrations down to ~5 p.p.m. where the MS detection limit for the product NOCl was approached. While this is approximately an order of magnitude higher than ambient, it is occasionally found in plumes from fossil fuel power plants. Because N₂O₄ concentrations are negligible at 5 p.p.m. NO₂, lowering the NO₂ concentration further to ambient should not alter the results reported.

Two types of experiments were performed: (1) IR studies at torr pressures of NO₂, in which simultaneous formation of NO₃⁻ and gaseous NOCl was observed and the stoichiometry $\Delta[\text{NO}_2]/\Delta[\text{NOCl}]$ measured, and (2) MS studies at mtorr concentrations of NO₂ (1 mtorr = 1.3 p.p.m., 25 °C and 1 atm) where NOCl was also identified and measured.

In the IR studies, 1.0–6.3 torr dry NO₂/N₂O₄ (prepared by oxidizing NO with O₂ and trapping the NO₂ at 196 K) was introduced into a gas cell (10 cm long × 3.2 cm diameter) with NaCl windows; the latter also provided the reacting surface. Initial concentrations of NO₂ and N₂O₄ were calculated from the measured total pressure using the published equilibrium constant¹⁴; their loss was monitored using the 1,750 cm⁻¹ band for N₂O₄ and the 1,630 and 2,930 cm⁻¹ bands for NO₂. NOCl was monitored at 1,790 cm⁻¹; its extinction coefficient was determined in separate experiments using authentic samples to be 2.9 × 10⁻² cm⁻¹ torr⁻¹.

Figure 1 shows a typical IR spectrum immediately after introducing 5.2 torr NO₂/N₂O₄ into the cell (a) and 96 min

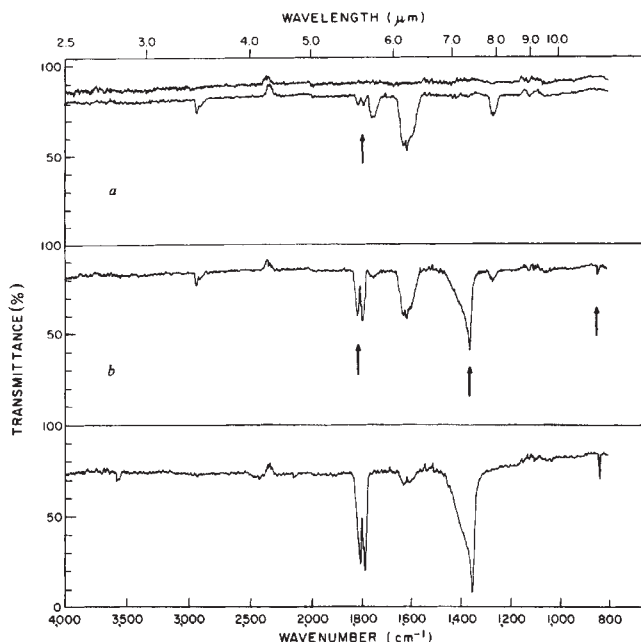


Fig. 1 IR spectra from the reaction NO₂ with NaCl at room temperature. Pressure of NO₂/N₂O₄ = 5.2 torr. a, Reaction time <10 min. b, Reaction time = 96 min. Arrows show growth of NOCl bands at ~1,800 cm⁻¹ and nitrate bands at 1,360 cm⁻¹ and 840 cm⁻¹. c, Reference NOCl spectrum at 3.1 torr. Bands at 1,360 cm⁻¹ and 840 cm⁻¹ are residual nitrate bands remaining on windows from previous NO₂/NaCl runs.

later (b). The reaction was sufficiently fast that NOCl (Fig. 1c) was observed even on the first scan. Nitrate bands at ~1,360 and 840 cm⁻¹ grew simultaneously. The formation of both NOCl and NO₃⁻ was initially linear with time, but eventually plateaued as expected for a surface reaction. The average stoichiometry $\Delta[\text{NO}_2]/\Delta[\text{NOCl}]$ in seven runs was 1.8 ± 0.3 (two standard deviations), consistent with reaction (1).

Having confirmed reaction (1) at torr reactant concentrations, a MS search for NOCl was carried out at mtorr pressures of NO₂ where the N₂O₄ concentration is negligible. Dilute (1%) NO₂/N₂O₄ mixtures in 1 atm of either He (ultra high purity, ≥99.995%) or laboratory air (passed through a glass wool-packed trap at 147 K) flowed through a needle valve into a flow system operated at 0.4 torr, giving an NO₂ pressure of 4 mtorr. A cell containing rock salt previously heated at 545 °C to remove organic impurities was located downstream of the needle valve. After flowing over the NaCl, the gas was sampled through a pinhole into a two-chamber quadrupole mass spectrometer where the beam was chopped at 800 Hz and the ions detected using phase sensitive detection. NOCl formation was monitored by the ³⁵Cl peak. The pressure of NOCl was estimated to be 1 × 10⁻² mtorr by comparison to ³⁵Cl peaks from measured NOCl concentrations. Identical results were obtained for both NO₂/air and NO₂/He mixtures and no ³⁵Cl was observed when NO₂ bypassed the salt. HCl was shown not to be the source of ³⁵Cl as a strong parent peak at m/e = 36 was not observed, as was the case for an authentic HCl sample.

Reaction (1) may be of particular importance in coastal urban areas due to the formation of NOCl. The potential for NOCl production can be estimated from the concentration of NaCl particles in marine environments, reported^{9,16–18} to be ~20–100 cm⁻³. Taking a typical particle radius^{16–18} of 1 μm and concentration of 30 cm⁻³, the total NaCl surface area available for reaction is ~4 × 10⁻¹⁰ m² per cm³ air. This corresponds to ~7 × 10⁹ molecules of NaCl on the surface if the area of one NaCl is assumed to be 5.8 Å². Oxidation of only the surface NaCl would thus result ultimately in a NOCl concentration of ~0.3 p.p.b. (1 p.p.b. = 2.5 × 10¹⁰ molecules cm⁻³ at 25 °C, 1 atm).

It is not known whether reaction (1) is sufficiently fast in typical ambient conditions to oxidize the entire surface. The fraction (ϕ) of NO_2 collisions with bulk artificial sea salt resulting in removal of NO_2 has been reported¹⁹ to be $\phi \approx 10^{-6}$ – 10^{-7} , the higher efficiencies occurring at higher relative humidities (RH). This is consistent with our IR studies from which we calculate $\phi \geq 5 \times 10^{-8}$ for a dry surface. The number of collisions per second (C) of NO_2 with the salt windows (exposed area A) was calculated from the usual gas kinetic expression $C = NA(RT/2\pi M)^{1/2}$. From C and the cell volume, the maximum possible increase in the NOCl concentration was calculated assuming that 0.5 NOCl was produced for each NO_2 collision. ϕ was then estimated as the ratio of the maximum experimentally observed rate of NOCl production to the maximum possible calculated rate. This approach is reasonable if a sequential type of reaction mechanism is operative, such as $\text{NO}_2 + \text{NaCl} \rightarrow \text{NO}_2\text{Cl}^- + \text{Na}^+$, $\text{NO}_2\text{Cl}^- + \text{NO}_2 \rightarrow \text{NO}_3^- + \text{NOCl}$. For 0.1 p.p.m. NO_2 , an NaCl surface area of $4 \times 10^{-10} \text{ m}^2$ per cm^3 of air, and $\phi = 10^{-6}$, ~ 0.1 p.p.b. NOCl , corresponding to $\sim 30\%$ of the surface NaCl , could be formed in 12 h (overnight, for example).

Three possible atmospheric fates of NOCl are: (1) reaction with OH , (2) photolysis, and (3) hydrolysis. The rate constant²⁰ for the NOCl – OH reaction at 25°C is $(4.3 \pm 0.4) \times 10^{-13} \text{ cm}^3 \text{ molecule}^{-1} \text{ s}^{-1}$. At a typical tropospheric OH concentration of $2 \times 10^6 \text{ cm}^{-3}$, the NOCl lifetime would be 14 days, too long to compete with photolysis and possibly hydrolysis.

NOCl strongly absorbs radiation in the actinic region, producing Cl atoms with a quantum yield of 1.0 (refs 21, 22)



Cl can initiate chain reactions characteristic of photochemical smog by reacting with organics to form HCl and a free radical in a manner similar to OH (ref. 1). Using published absorption cross-sections^{23,24} and the actinic irradiance²⁵, the NOCl photolytic lifetime is 5 min at a zenith angle (z) of 0° , and 27 min at $z = 80^\circ$. For $z = 80^\circ$ (that is, early morning), the rate of Cl atom production is estimated to be ~ 2 – $6 \times 10^6 \text{ atoms cm}^{-3} \text{ s}^{-1}$ for 0.1–0.3 p.p.b. of NOCl . This can be compared with a rate of OH production of $\sim 4 \times 10^7 \text{ radicals cm}^{-3} \text{ s}^{-1}$ from 3 p.p.b. of HONO , typical of polluted urban atmospheres²⁶. Thus photolysis of NOCl even at sub-p.p.b. concentrations could account for $\sim 10\%$ of organic oxidations in coastal atmospheres, and possibly more if ϕ is found in more detailed studies to be greater than 10^{-6} .

Hydrolysis of NOCl likely produces HONO , a major photolytic source of OH in polluted urban atmospheres¹:



While hydrolysis of NOCl in aqueous solutions is thought to be fast, a gas phase rate constant has not been reported. If it is in the range 10^{-16} – $10^{-21} \text{ cm}^3 \text{ per molecule per s}$, then at 50% RH and 25°C , the lifetime with respect to reaction (3) is 0.03 s–44 min, competitive with photolysis. Reaction (1) followed by reaction (3) is consistent with the observation^{11,12} of NaNO_3 and HCl in the NO_2 – NaCl reaction.

Reaction of NO_2 with Cl^- in the liquid phase (for which there is some evidence at high concentrations²⁷) should also be considered, because NaCl is hygroscopic and dissolves at $\text{RH} \geq 75\%$. Production of NOCl in aqueous droplets would provide a source of HONO and possibly chlorine atoms in these aerosols, which could then have a role in aerosol chemistry associated with acid deposition, as well as SO_2 and hydrocarbon oxidations.

While only gross estimates of the atmospheric NOCl production by reaction (1) and the fate of NOCl are possible due to the kinetic uncertainties, NOCl could clearly contribute to coastal urban atmospheric chemistry. Further studies are underway in this laboratory to establish the kinetics of reaction (1), as well as of the relevant NOCl reactions.

I thank Dr G. J. Doyle, Dr R. Atkinson and Dr P. A. Wegner for helpful discussions, the NSF for IR and MS equipment grants

(numbers TFI-8020406 and PRM-8206043) and California State University Fullerton for a Faculty Research Grant.

Received 31 May; accepted 6 October 1983.

1. Finlayson-Pitts, B. J. & Pitts, J. N. Jr *Adv. Envir. Sci. Technol.* **7**, 75–162 (1977).
2. Calvert, J. G. (ed.) *Acid Precipitation: SO_2 , NO and NO_2 Oxidation Mechanisms. Atmospheric Considerations* (Ann Arbor Science, 1983).
3. Waldman, J. M. et al. *Science* **218**, 677–680 (1982).
4. Pitts, J. N. Jr *Envir. Hlth Perspect.* **47**, 115–140 (1983).
5. Altschuler, A. P. *Tellus* **10**, 479–492 (1958).
6. Schroeder, W. H. & Urone, P. *Envir. Sci. Technol.* **8**, 756–758 (1974).
7. Junge, C. E. *Tellus* **8**, 127–139 (1956).
8. Green, W. D. *J. geophys. Res.* **77**, 5152 (1972).
9. Hidy, G. M. et al. *J. appl. Met.* **13**, 96–107 (1974).
10. Kobayashi, T., Ikezawa, T. & Watanabe, H. *Taiiki Osen Gakkaishi* **14**, 401–407 (1979).
11. Robbins, R. C., Cadle, R. D. & Eckhardt, D. L. *J. Met.* **16**, 53–56 (1959).
12. Cadle, R. D. & Robbins, R. C. *Disc. Faraday Soc.* **30**, 155–161 (1960).
13. Chung, T. T., Dash, J. & O'Brien, R. J. *9th Int. Cong. Electron. Microsc.* 440–441 (1978).
14. Stull, D. R. & Proffitt, H. (eds) *JANAF Thermochemical Tables*, 2nd edn (NSRDS-NBS 37 US Department of Commerce, National Bureau of Standards, 1971).
15. Breuer, H. D. & Krüger, J. *Ber. Bunsenges. Phys. Chem.* **82**, 97 (1978).
16. Junge, C. E. *J. geophys. Res.* **77**, 5183–5200 (1972).
17. Blanchard, D. C. & Woodcock, A. H. *Ann. N. Y. Acad. Sci.* 330–347 (1980).
18. Patterson, E. M. et al. *J. geophys. Res.* **85C**, 7361–7376 (1980).
19. Sverdrup, G. M. & Kuhlman, M. R. *Stud. Envir. Sci.* **8**, 245–248 (1980).
20. Poulet, G., Jourdain, J. L., Laverdet, G. & Le Bras, G. *Chem. phys. Lett.* **81**, 573–577 (1981).
21. Calvert, J. G. & Pitts, J. N. Jr *Photochemistry* (Wiley, New York, 1966).
22. Grimley, A. J. & Houston, P. L. *J. chem. Phys.* **72**, 1471–1475 (1980).
23. Hampson, R. F. *Chemical Kinetic and Photochemical Data Sheets for Atmospheric Reactions* (Rep. no. FAA-EE-80-17, US Department of Transportation, 1980).
24. Ballash, N. M. & Armstrong, D. A. *Spectrochim. Acta* **30A**, 941–944 (1974).
25. Leighton, P. A. *Photochemistry of Air Pollution* (Academic, New York, 1961).
26. Harris, G. W. et al. *Envir. Sci. Technol.* **16**, 414–419 (1982).
27. Lilly, E. I. & Co., French Pat. 1,556,031 (1969); *Chem. Abstr.* **8th Coll. Index** **66**–75, 1967–1971.

P-wave travel time and amplitude in western North America

Rhett Butler

Hawaii Institute of Geophysics, University of Hawaii, Honolulu, Hawaii 96822, USA

A study of teleseismic, short-period (1 s) P-wave travel-time and amplitude variations in western North America reported here reveals a high level of correlation: early arrivals relate to low amplitudes and late arrivals to higher amplitudes. This trend is opposite to observations noted for the USA and southern Canada as a whole¹. With some exceptions, seismic stations east of the Rocky Mountain front generally record faster travel times^{2–8} and higher amplitudes^{9–15} than those to the west. These differences have been attributed to higher attenuation and lower velocities in the upper mantle beneath the tectonically active western region^{11,16}. The correlation in the west between P-wave travel-time residuals and relative amplitudes is consistent with elastic focusing and defocusing effects and is found for both World Wide Standardized Seismograph Network (WWSSN) and Long Range Seismic Measurement (LRSM) stations. These results are relevant to the estimation of yields of nuclear explosions because by reciprocity similar elastic focusing and defocusing effects can alter the amplitudes and travel times of P-waves from explosions.

P-wave travel-time residuals for WWSSN^{3,4,6–8} and LRSM^{2–4} studies were reduced to a common baseline by using International Seismological Center (ISC) travel times⁵. Then each station residual was computed by the maximum likelihood method of averaging time residuals weighted by the reciprocals of their variances. P-wave amplitudes are relative to their average value over the USA and southern Canada¹⁵. Mean $\log_{10}$ of amplitude are averaged for WWSSN over azimuth from Butler¹⁵ and for LRSM stations from three LRSM studies^{10–12}. For western North America a least-squares line, $y = ax + b$, is

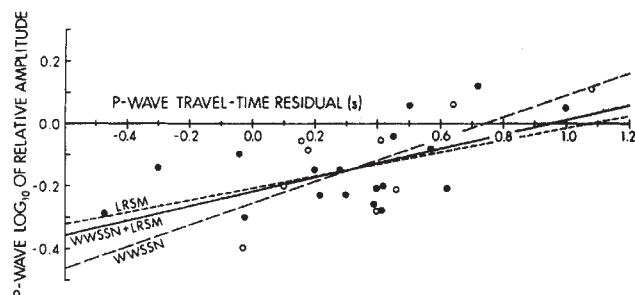


Fig. 1 Relationship between P-wave travel-time residual and $\log_{10}$ of relative amplitude for seismic stations in western North America. World Wide Standardized Seismograph Network (WWSSN) and Long Range Seismic Measurement (LRSM) stations are plotted as open and closed circles, respectively. Positive travel times are slow and positive $\log_{10}$ of amplitude is larger than average. Least-squares lines are plotted for WWSSN, LRSM and joint WWSSN-LRSM data sets.

fitted, where x = P-wave travel-time residual (s) and $y = \log_{10}$ of relative amplitude.

y ($= \log_{10}$ of relative amplitude) data are plotted as a function of x (P-wave travel-time residuals) in Fig. 1 for the WWSSN and LRSM stations. The least-squares lines, $y = (0.349)x - (0.256)$ for the WWSSN and $y = (0.194)x - (0.208)$ for the LRSM data, are similar. The correlation (r) between x (P-wave travel-time residuals) and y ($\log_{10}$ of relative amplitude) is significant at a 95% level of confidence for both the WWSSN ($r = 0.716$) and LRSM ($r = 0.545$) data for their number of degrees of freedom¹⁷. The least-squares line for the joint WWSSN-LRSM data is $y = (0.241)x - (0.220)$ and the correlation ($r = 0.604$) between x and y is significant at a 99% level of confidence for the total number of degrees of freedom.

The positive correlation (or slope) between P-wave travel-time residuals and $\log_{10}$ of relative amplitudes in Fig. 1 for seismic stations in western North America is in stark contrast to the negative correlation found for the USA and southern Canada as a whole, and two different controlling mechanisms are suggested. On the large scale, the variations between western and eastern North America are probably rooted in lateral differences in temperature. Higher temperatures beneath the tectonically active west produce higher attenuation of P-waves, lower velocities in the upper mantle, and high surface heat flow. Within the western province the P-wave travel times are correlated with amplitudes in a way qualitatively explained by elastic focusing and defocusing. A low-velocity lens near a seismic station can slow travel times while focusing the arriving energy; a high-velocity lens can yield faster times while defocusing the arriving energy. Although the image of a lens is used to illustrate the elastic focusing and defocusing effect, the anomalous region near a seismic station could also include undulations in the depth to velocity discontinuities, and random distributions of high- and low-velocity bodies. The size of the anomalous region could range from a single P wavelength (~ 6 – 8 km for a wave with a 1-s period) to many wavelengths, but large regions will affect a number of seismic stations. The positive correlation between P-wave travel time and $\log_{10}$ of amplitude has been reported as a characteristic of the large seismic arrays LASA^{18–20} in Montana, USA, and NORSAR^{19,21} in Norway. These studies^{18–21} have attributed a large part of their amplitude and travel-time variations to elastic focusing and defocusing resulting from lateral velocity heterogeneity in the upper 200 km, both in terms of random^{18,19} and deterministic structure^{20,21}. Thus, over the area of seismic arrays 10^4 km², elastic focusing and defocusing have an important role. The results presented here extend the important role of that mechanism to western North America over an area several times 10^6 km² in size. Further-

more, whereas subarray travel-time and amplitude variations within LASA and NORSAR may be caused by underlying structure common to the subarrays, the P-wave travel-time and amplitude variations associated with seismic stations in western North America are most likely to be independent between stations and not caused by common structure. The scatter about the least-squares lines in Fig. 1 may be due to sub-regional variations of anelastic attenuation and velocity, signal amplification by low-velocity sedimentary layers, and, perhaps, topographic effects.

A similar study of P-wave travel time and amplitude relations east of the Rocky Mountains does not show significant correlations. However, large areas of soft sediments covering much of central and eastern North America effectively amplify the seismic signals at stations without substantially changing the P-wave travel time¹⁴. Yet, seismic stations in eastern North America situated on hard bedrock show a negative x, y correlation ($r = -0.318$) with fast P-wave times relating to large amplitudes—as is the case with the United States and southern Canada as a whole—but not significant at a 95% confidence level. A recent analysis²² of S-wave relative amplitudes and travel-time residuals of the WWSSN and Canadian network has noted that, in general, western North America records lower amplitudes and later arrivals relative to the east—the negative correlation also noted for P waves¹. However, within western North America the parameters evinced a positive correlation for both short-period ($r = 0.428$) and long-period ($r = 0.381$) waves. For their number of degrees of freedom, these correlations are significant only slightly below the 95% level of confidence. In eastern North America a positive correlation was also observed²², although at a low level of confidence, between S-wave travel-time residuals and amplitudes. Both P-waves and S-waves evince a positive correlation between their respective travel-time and amplitude variations in western North America. In the east, although the trends between P-wave and S-wave travel time and amplitudes are opposite, their statistical significance is not high.

I conclude that the correlation in western North America between travel-time residuals and relative amplitudes is qualitatively consistent with elastic focusing and defocusing; slow times relate to higher amplitudes and fast times to low amplitudes. This relationship is observed for both P-waves and S-waves and at WWSSN, LRSM, and Canadian network seismic stations. In eastern North America this systematic relationship is not observed. The cause underlying this dichotomy between western and eastern North America is not clear. Possibly, the tectonically active west has a manifestly greater variation of upper mantle and crustal velocity. Alternatively, amplitude variation in the east may be primarily controlled by attenuation variation. Hawaii Institute of Geophysics contribution no. 1386.

Received 31 May; accepted 10 October 1983.

- Der, Z. A., Masse, R. P. & Gurski, J. P. *Geophys. J. R. astr. Soc.* **40**, 85–106 (1975).
- Cleary, J. R. & Hales, A. L. *Bull. seism. Soc. Am.* **56**, 467–489 (1966).
- Carder, D. S., Gordon, D. W. & Jordan, J. N. *Bull. seism. Soc. Am.* **56**, 815–840 (1966).
- Herrin, E. & Taggart, J. *Bull. seism. Soc. Am.* **58**, 1325–1337 (1968).
- Evernden, J. & Clark, D. M. *Phys. Earth planet. Inter.* **4**, 1–23 (1970).
- Lilwall, R. C. & Douglas, A. *Geophys. J. R. astr. Soc.* **19**, 165–181 (1970).
- Sengupta, M. K. & Julian, B. R. *Bull. seism. Soc. Am.* **66**, 1555–1579 (1976).
- Romanowicz, B. *Geophys. J. R. astr. Soc.* **57**, 479–506 (1978).
- Jordan, J., Black, R. & Bates, C. C. *Bull. seism. Soc. Am.* **55**, 693–720 (1965).
- Cleary, J. J. *geophys. Res.* **72**, 4705–4712 (1967).
- Evernden, J. & Clark, D. M. *Phys. Earth planet. Inter.* **4**, 24–31 (1970).
- Booth, D. C., Marshall, P. D. & Young, J. B. *Geophys. J. R. astr. Soc.* **39**, 523–537 (1974).
- North, R. G. *Lincoln Lab. Tech. Note TN-1977-24* (MIT, Cambridge, 1977).
- Butler, R. & Ruff, L. *Bull. seism. Soc. Am.* **70**, 831–850 (1980).
- Butler, R. *Rev. Geophys. Space Phys.* (in the press).
- Hales, A. L. & Herrin, E. in *The Nature of the Solid Earth* (ed. Robertson E.) (McGraw-Hill, New York, 1977).
- Snedecor, G. W. & Cochran, W. G. *Statistical Methods* (Iowa State University Press, 1967).
- Aki, K. J. *geophys. Res.* **78**, 1334–1346 (1973).
- Berteussen, K. A., Christofferson, A., Husebye, E. S. & Dahle, A. *Geophys. J. R. astr. Soc.* **42**, 403–417 (1975).
- Haddon, R. A. & Husebye, E. S. *Geophys. J. R. astr. Soc.* **55**, 19–43 (1978).
- Chang, A. C. & Von Seggern, D. H. J. *geophys. Res.* **85**, 4811–4828 (1980).
- Lay, T. & Helmberger, D. V. *Bull. seism. Soc. Am.* **73**, 1063–1076 (1983).

Isotopic evidence from the eastern Canadian shield for geochemical discontinuity in the Proterozoic mantle

L. D. Ashwal* & J. L. Wooden*

Lunar and Planetary Institute, 3303 NASA Road 1, Houston, Texas 77058, USA

Most workers agree that Proterozoic anorthosite massifs represent the crystallization products of mantle-derived magmas^{1,2}, although the composition of the parental melts is a major unsolved petrological problem³. As mantle-derived rocks, the massifs can be used as geochemical probes of their late Precambrian upper mantle sources. We report here Nd and Sr isotopic compositions of anorthosites and related rocks from the Grenville and Nain Provinces of the eastern Canadian shield. Here 75% of the Earth's known anorthosite is found in a 1,600-km belt from the Adirondack Mountains of northern New York State to the eastern coast of Labrador⁴ (Fig. 1). The results indicate that the massifs were derived from at least two distinct mantle source regions which were established before 1,650 Myr ago, and were episodically involved in magmatism over ~500 Myr. One reservoir, below the Grenville Province, and probably below much of the eastern Superior Province, was isotopically similar to the depleted, modern-day mid-ocean ridge basalt (MORB) source. The other reservoir was chondritic to moderately enriched, and is most easily identified in the Nain Province, but may have occurred scattered throughout the Superior Province.

The anorthosites were emplaced into presumed Archaean quartzofeldspathic gneisses, although in large areas of the Grenville Province, both massifs and basement rocks were reworked during late Proterozoic (about 1,000 Myr) Grenvillian metamorphism⁵. The Adirondack anorthosite may have crystallized as much as 300 Myr before high-pressure granulite facies metamorphism⁵. The other anorthosite massifs that we analysed are unmetamorphosed and include the large (7,000–8,000 km²) composite massifs at Harp Lake⁶ and Mealy Mountains, Labrador⁷, and a smaller complex (300 km²) at St Urbain, Quebec⁸ (Fig. 1). Isotopic data have been published for three mafic intrusions in Labrador: the Kiglapait layered intrusion^{9–11} and the Flowers River gabbro¹², both associated with the Nain anorthosite complex, and the Shabogamo gabbro, which intrudes supracrustal rocks of the Labrador Trough on both sides of, but within 25 km of the Grenville Front^{13,14} (Fig. 1).

Sr and Nd isotopic data for these anorthositic and mafic intrusions are given in Table 1. The massifs can be divided into two types based on location and Nd isotopic signature (Fig. 2). This division is possible despite uncertainties of up to 200 Myr in age and five epsilon units in initial Nd isotopic ratio (ϵ_{Nd} as defined in ref. 15), caused by the effects of metamorphism and crustal contamination⁵. South-east of a line roughly parallel to the Grenville Front (Fig. 1) anorthosites have positive ϵ_{Nd} values, implying sources with long-term depletion in light rare earth elements (LREE). To the north-west, anorthosites and mafic intrusions in Labrador have negative ϵ_{Nd} values that suggest derivation from LREE-enriched sources. With the exception of the Adirondack anorthosite, initial Sr isotopic ratios (Sr_0) also permit discrimination between enriched ($Sr_0 \geq 0.704$) and depleted ($Sr_0 \leq 0.703$) sources (Table 1). The enriched Sr_0 of the Adirondack anorthosite may be related to its anomalously high $^{18}O/^{16}O$, which has been attributed to isotopic exchange with pervasive metamorphic fluids¹⁶, or to bulk assimilation of metasedimentary units by the anorthosite parental magma¹⁷.

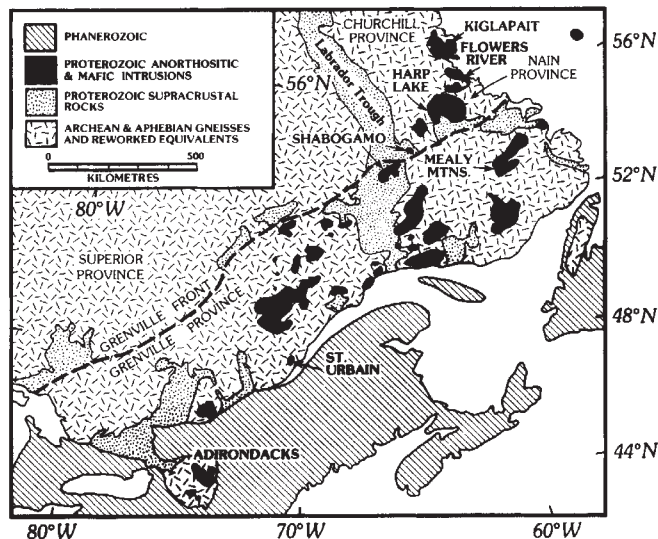


Fig. 1 Geological map of eastern North America showing locations of major Proterozoic anorthosite massifs and mafic intrusions.

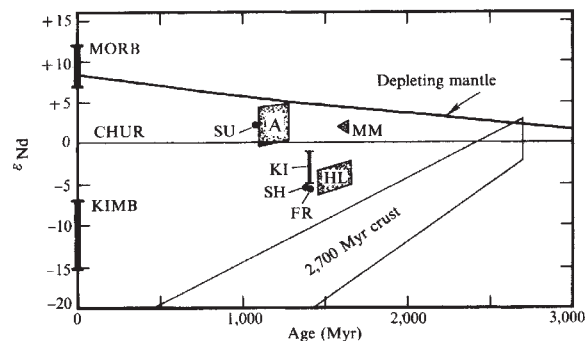


Fig. 2 Plot of age versus ϵ_{Nd} for Proterozoic anorthositic and mafic intrusions from the eastern Canadian shield. Massifs derived from sources with a depleted signature plot above the chondritic evolution line (CHUR), and include St Urbain (SU), Adirondacks (A) and Mealy Mountains (MM). Intrusions plotting below the CHUR line including Kiglapait (KI), Shabogamo (SH), Flowers River (FR) and Harp Lake (HL) may have been derived from enriched reservoirs. Data for some massifs are shown as polygons because of age uncertainties and contamination effects. Curve representing evolution of depleting mantle is from ref. 15; field showing evolution of 2,700-Myr silicic crust is from unpublished data of J.L.W. Ranges of ϵ_{Nd} for modern-day MORB²⁹ and Australian kimberlites (KIMB)¹⁸ are also shown.

It must be considered plausible that the Labrador massifs north-west of the Grenville Front do not reflect the isotopic compositions of their mantle sources, but instead were affected by a 'crustal' contaminant enriched in $^{87}Sr/^{86}Sr$ and depleted in $^{143}Nd/^{144}Nd$. It has been suggested, for example, that at least 20% assimilation of a granitic component by a mantle-derived mafic melt would be required to lower the ϵ_{Nd} of the Shabogamo gabbro from 0 to -5.3 (ref. 13). Much larger volumes of contaminant would be required if the primary melts were originally derived from a depleted source (Fig. 2). It is not clear that sufficient chemical, petrographic or field evidence exists to corroborate such extensive contamination for the Shabogamo gabbro or the other Labrador massifs. In fact, Morse¹¹ presents arguments against such effects for the Kiglapait intrusion. Selective contamination models involving fluids, for example, are more plausible, but there is no reason why such effects should be limited to the terrain north of the Grenville Front.

We think it is more likely that the Proterozoic massifs were derived from two distinct reservoirs: one was a depleted source, similar to the modern-day MORB source, the other was either a moderately enriched source, or a roughly chondritic one if

* Present addresses: Department of Geology and Geophysics, Yale University, New Haven, Connecticut 06520, USA (L.D.A.); US Geological Survey, 345 Middlefield Road, Menlo Park, California, USA (J.L.W.).

Table 1 Isotopic compositions of Proterozoic anorthosites and mafic intrusions

	Age (Myr)	Sr ₀	ε _{Nd}	Ref.
Depleted source				
Adirondacks, New York	1,288 ± 36*	0.7039–0.7048	+0.4 to +4.1	5, 26
	1,113 ± 10†	0.7040–0.7055	–0.3 to +4.4	
St Urbain, Quebec	1,079 ± 22*	0.7029	+2.2	8, this work
Mealy Mt., Labrador	1,576*		+2.0	7, this work
	1,640 ± 3†	0.7026–0.7032	+1.3 to +2.8	
Enriched (?) source				
Harp Lake, Labrador	1,663 ± 45*	0.7040–0.7066	–5.3 to –2.0	27, 28, this work
	1,460†	0.7040–0.7068	–6.1 to –3.4	
Kiglapait, Labrador	1,416 ± 50*	0.7039–0.7066	–4.9 to –1.1	9–11
Shabogamo, Labrador	1,379 ± 65*	0.7041	–5.3	13, 14
Flowers River, Labrador	1,411 ± 48*		–5.6	12
	1,200 ± 210‡	0.7065		

* From Sm–Nd whole-rock and mineral isochron.

† U–Pb zircon age of cross-cutting granitic intrusion.

‡ Rb–Sr errorchron.

contamination has had a role in altering the isotopic compositions of the Labrador massifs. The possibility of having identified a mid-Proterozoic enriched source is particularly interesting in view of the accumulating evidence for enriched reservoirs in present-day sub-continental settings^{18–21}. Both depleted and enriched (?) sources had to be juxtaposed before 1,600–1,650 Myr ago to give rise to the Mealy Mountains and Harp Lake massifs (Table 1). However, both sources must have been formed much earlier, particularly if they developed from a single reservoir with an originally chondritic Sm/Nd. Determination of the ages and extent of these reservoirs beneath the North American continent awaits further isotopic study.

Evidence for a regionally-depleted mantle beneath the Superior Province of North America throughout the Proterozoic comes from Sr and Pb isotopic data for alkaline complexes in Ontario^{22,23}. There is also a growing body of Nd isotopic data for the late Archaean rocks of the Superior Province that indicates mantle sources with ε_{Nd} of +1 to +4.5 (ref. 24). A depleted mantle signature is also recorded in middle Proterozoic rocks of the western USA¹⁵. Therefore, it would almost be expected that the mantle-derived middle Proterozoic rocks of the Grenville Province should have positive ε_{Nd} values. In this context, it is the Nain Province rocks, with their enriched signature, that are unusual. If the isotopic signature of these rocks truly reflects enriched mantle, and not crustal contamination, then this enriched reservoir might consist of: (1) small mantle regions beneath North America that did not participate in Archaean crustal genesis, (2) depleted mantle that was enriched by metasomatic processes shortly after the major Archaean crust formation event or (3) a tapping of a primitive, undepleted source from a deep mantle reservoir²⁵.

We thank H. Wisemann, C.-Y. Shih and B. Bansal for analytical assistance, L. E. Nyquist for access to the JSC mass spectrometry laboratory, and R. F. Emslie, L. P. Gromet, R. F. Dymek, and H. W. Jaffe for providing samples. LPI contribution no. 521.

Received 24 June; accepted 4 October 1983.

- Morse, S. A. *Am. Miner.* **67**, 1087–1100 (1982).
- Emslie, R. F. *Precamb. Res.* **7**, 61–98 (1978).
- Ashwal, L. D. *Lunar planet. Inst. Tech. Rep.* **82-01**, 40–44 (1982).
- Wynne-Edwards, H. R. *Geol. Ass. Can. Spec. Pap.* **11**, 264–334 (1972).
- Ashwal, L. D. & Wooden, J. L. *Geochim. cosmochim. Acta* (in the press).
- Emslie, R. F. *Geol. Surv. Can. Bull.* **293**, (1980).
- Emslie, R. F., Loveridge, W. D., Stevens, R. D. & Sullivan, R. W. *Geol. Ass. Can.-Miner. Ass. Can. Prog. Abstr.* **8**, A-20 (1983).
- Gromet, L. P. & Dymek, R. F. *Lunar planet. Inst. Tech. Rep.* **82-01**, 72–74 (1982).
- DePaolo, D. J. *Geol. Soc. Am. Abstr. Prog.* **13**, 437 (1981).
- Simmons, E. C. & Lambert, D. D. *Univ. Mass. Dep. Geol. Contr.* **38**, 41–46 (1981).
- Morse, S. A. *Science* **220**, 193–195 (1983).
- Brooks, C. in Hill, J. D. *Newfoundland Dep. Mines Energy Rep.* **82-6**, (1982).
- Zindler, A., Hart, S. R. & Brooks, C. *Earth planet. Sci. Lett.* **54**, 217–235 (1981).
- Brooks, C., Wardle, R. J. & Rivers, T. *Can. J. Earth Sci.* **18**, 1211–1227 (1981).
- DePaolo, D. J. *Nature* **291**, 193–196 (1981).
- Taylor, H. P. *Jr N.Y. State Mus. Sci. Serv. Mem.* **18**, 111–134 (1969).
- Valley, J. W. & O'Neil, J. R. *Nature* **300**, 497–500 (1982).
- McCulloch, M. T., Jaques, A. L., Nelson, D. R. & Lewis, J. D. *Nature* **302**, 400–403 (1983).

- Menzies, M. & Murthy, V. R. *Nature* **283**, 634–636 (1981).
- Collerson, K. D. & McCulloch, M. T. *Proc. 4th. int. Symp. Antarctic Earth Sci.* 676–680 (Cambridge University Press, 1983).
- Erlank, A. J., Allsopp, H. L., Hawkesworth, C. J. & Menzies, M. A. *Terra Cognita* **2**, 261–263 (1982).
- Bell, K., Blenkinsop, J., Cole, T. J. S. & Menagh, D. P. *Nature* **298**, 251–253 (1982).
- Tilton, G. R. *Abstr. for 1983 Archaean Geochem.-Early Crustal Genesis Field Wkshp* (in the press).
- Wooden, J. L., Bansal, B. & Wiesmann, H. *EOS* **64**, 340 (1983).
- Wasserburg, G. J. & DePaolo, D. J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 3594–3598 (1979).
- Silver, L. T. *N.Y. State Mus. Sci. Serv. Mem.* **18**, 233–251 (1969).
- Wooden, J. L., Emslie, R. F. & Ashwal, L. D. *Geol. Soc. Am. Abstr. Prog.* **14**, 649 (1982).
- Krogh, T. E. & Davis, G. L. *Yb Carnegie Inst Wash.* **72**, 610–613 (1973).
- Bas. Volc. Stud. Proj. *Basaltic Volcanism on the Terrestrial Planets* (Pergamon, New York, 1981).

Post-glacial arid episodes in Ethiopia have implications for climate prediction

Richard Gillespie*, F. Alayne Street-Perrott† & Roy Switsur‡

* Tropical Palaeoenvironments Research Group, and Radiocarbon Accelerator Unit, Research Laboratory for Archaeology, 6 Keble Road, Oxford OX1 3QJ, UK

† Tropical Palaeoenvironments Research Group, and School of Geography, Mansfield Road, Oxford OX1 3TB, UK

‡ The Godwin Laboratory, Sub-Department of Quaternary Research, Free School Lane, Cambridge CB2 3RS, UK

Several lines of geomorphological and stratigraphical evidence suggest that the widespread early post-glacial wet phase in tropical Africa between 12,500 and 5,000 yr BP (refs 1, 2) was interrupted by severe dry episodes during the intervals 11,000–10,000, 8,500–6,500 and 6,200–5,800 yr BP (refs 3–9). The magnitude of the second event has recently been questioned¹⁰. We present new stratigraphical and dating evidence from the Ziway-Shala lake basin, Ethiopia, indicating the severity and timing of dry phases during the interval 12,000–5,000 yr BP. Our findings have important implications for predictions of the climatic impact of increasing atmospheric levels of CO₂.

The Ziway-Shala Basin (7°–8°30' N) is a closed drainage in the Main Ethiopian Rift. The four modern lakes are the shrunken remnants of a large late Quaternary palaeolake (2,960 km²), with an outlet at ~1,670 m above sea level (a.s.l.) into the adjacent Awash catchment (Fig. 1)^{11,12}. The lowest lake, Shala, is 266 m deep. Its surface stood at ~1,558 m in 1972, which is referred to as Shala datum (SD).

Fluctuations in water level during the late Quaternary have been reconstructed from several types of evidence: dated shorelines, stratigraphical studies of the rift-floor sediments¹², and analyses of the pollen¹³, diatoms¹⁴ and chemistry (F.A.S-P.,

Table 1 Stratigraphical and dating evidence for low levels of Lake Ziway-Shala, 14,000–0 yr BP

Event	Area	Facies	Lake Shala level (m SD)	¹⁴ C age (yr BP)	Lab. no.	Material dated	Refs
Ziway-Shala VI/VII	A	Alluvial silt	<42	2,510 ± 100	Gif-4011	<i>Acacia</i> charcoal	
Minor recession during phase	A	Thin bed of reworked shells with articulated	?	5,530 ± 65	Q-1369	Mixed shell	
Ziway-Shala VI		fish skeletons (ref. 12)		5,680 ± 55	Q-1557	Mixed shell	
				5,750 ± 70	Q-1361	Fishbone collagen	
				5,980 ± 100	GaK-3387	<i>Melanoides</i> shell	11
				6,495 ± 150	SUA-497	<i>Melanoides</i> shell	
Ziway-Shala V/VI	D	Colluvium	<76	>> 5,940 ± 70	Q-1359	See Table 2	
				<< 9,270 ± 85	Q-1358	See Table 2	
	C	Palaeosol	<61	8,520 ± 200	Gif-3984	Soil organic matter	
	B	Shallow-water lacustrine sands with salt-tolerant diatoms	<25	<7,970 ± 150	Gif-3965	Lake mud	12–14
	A	Ephemeral-stream gravels	<3	>> (5,530–6,495)		See above	
				<< 8,655 ± 175	Q-1114	See Table 2	
Ziway-Shala IV/V	C	Ephemeral-stream gravels	<57	>> 8,520 ± 200	Gif-3984	See above	
				<< 11,870 ± 300	SUA-494	See below	
	B	Palaeosol, overlying	<22	9,950 ± 170	Gif-3968	Soil organic matter	12–14
		Shallow-water lacustrine sands with reworked diatoms	<19	<10,370 ± 180	Gif-3969	Lake mud	12–14
	A	Exhumed <i>Acacia</i> stumps	<–1	9,800 ± 100	HAR-2787	Carbonized wood	
		Palaeosol	<71	10,220 ± 140	SUA-500	<i>Acacia</i> charcoal	
Ziway-Shala III/IV	C	Palaeosol	<57	10,450 ± 180*	Gif-3986	Soil organic matter	29
				11,870 ± 300	SUA-494	Charcoal	30
		Palaeochannel fill	<44	12,230 ± 150	SUA-504	<i>Caecatur</i> shell	30

All ¹⁴C dates have been corrected for isotopic fractionation using ¹³C measurements if available or group mean values for similar materials. Areas (Fig. 1): A, Ajewa embayment; B, Abiyata Core; C, Dekka Wede type section; D, Boloko section.

* Stratigraphically inconsistent by >2σ.

unpublished data) of a long core from Lake Abiyata (Fig. 1B). The lake-level sequence for the past 14,000 yr is based on 40 ¹⁴C dates (Tables 1, 2).

Deglaciation of the mountains on the eastern watershed was completed before 11,500 yr BP (refs 12, 15). The lakes rose around 12,000 yr BP after a prolonged low stand (Fig. 2). During phase Ziway-Shala IV, they united at ~+74 m SD. A major recession had set in by ~10,400 yr BP, and is recorded by dated palaeosols in the Ajewa embayment (10,220 ± 140 yr BP) and the Abiyata core (9,950 ± 170 yr BP) (Table 1). A date of 9,800 ± 100 yr BP from *Acacia* stumps exposed on the eastern shore of Lake Shala (Fig. 1A) demonstrates that the lake receded to at least 1 m below SD. The subsequent transgression took place so rapidly that there is considerable overlap between ¹⁴C dates on underlying, nonlacustrine and overlying, lacustrine materials (Fig. 2).

During phase Ziway-Shala V, the united lake rose to at least +102 m SD, the altitude of the bedrock lip underlying the sedimentary fill in its spillway. It may, therefore, have overflowed during this phase. Around 8,500 yr BP, the water level began to fall again, reaching +3 m SD or even lower. Ephemeral-stream gravels, deposited during this regression near the present eastern shore of Lake Shala, are overlain by lacustrine silts which have yielded dates ranging from 6,495 to 5,530 yr BP (Table 1). In the Boloko section (Fig. 1D), the Ziway-Shala V/VI recession is represented by 3.7 m of colluvial deposits. These are rather widely bracketed by ¹⁴C dates of 9,270 ± 85 and 5,940 ± 70 yr BP from the adjacent lacustrine beds. Based on all the available stratigraphical and dating evidence, the lowest levels occurred around 7,800–7,200 yr BP (Fig. 2).

The succeeding transgression (VI) culminated in an overflow of the united lake at +112 m SD. It was interrupted by a short recession centred on 5,900 yr BP (Table 1), which caused a massive killing of fish in the Shala Basin, implying a catastrophic breakdown of stratification¹². This abrupt drop in lake level probably lasted a few centuries at most. Its amplitude is uncertain. A measured age of 5,180 ± 55 yr BP (Table 2) from a beach at +110.5 m SD on the south-west side of Lake Langan, marks

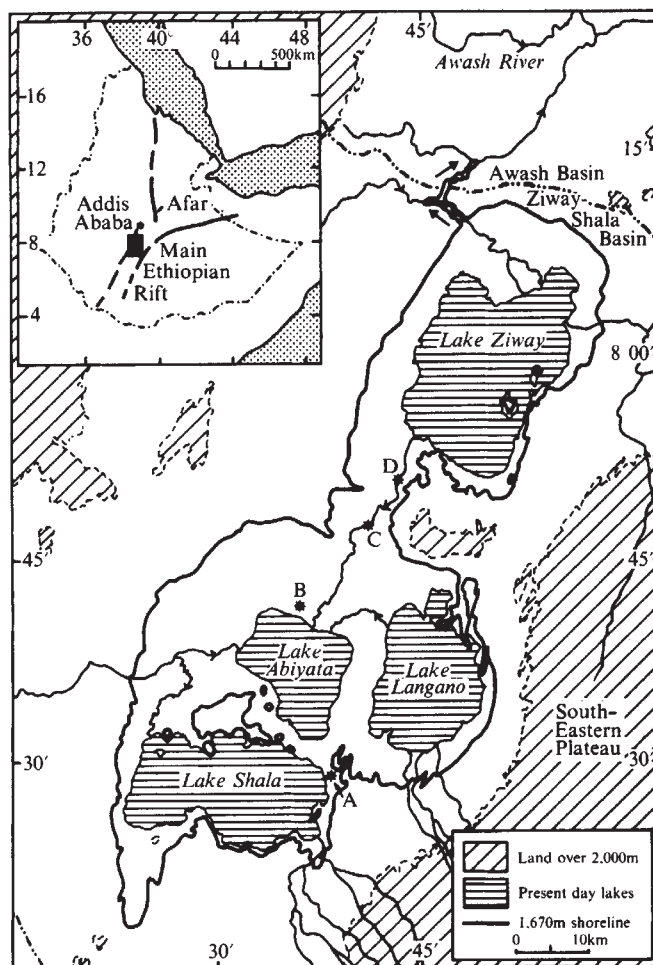


Fig. 1 Location map of palaeolake Ziway-Shala. A, Ajewa embayment; B, Abiyata Core; C, Dekka Wede type section; D, Boloko section.

Table 2 ^{14}C dates recording high levels of Lake Ziway-Shala, 14,000–0 yr BP

Event	Facies	^{14}C age (yr BP)	Lab no.	Material dated	Elevation of sample (m SD)	Refs
Ziway-Shala VI	Shore terrace	5,160 $\pm$ 75	SMU-65	<i>Bulinus</i> shell	~106	32
	Beach sands	5,180 $\pm$ 55	Q-1360	Mixed shell	110.5	
	Coquina*	5,320 $\pm$ 145	Not publ.	Shell	87‡	31
	Shore terrace	5,410 $\pm$ 105	SMU-67	<i>Bulinus</i> shell	~106	32
	Coquina	5,660 $\pm$ 125	Gif-4014	<i>Melanoides</i> shell	55‡	
	Shore terrace	5,730 $\pm$ 65	SMU-69	<i>Bulinus</i> shell	~95	32
	Coquina	5,810 $\pm$ 125	Gif-4012	<i>Bellamya</i> shell	55‡	
	Coquina	5,910 $\pm$ 125	Gif-4013	<i>Bulinus</i> shell	55‡	
	Shore terrace	5,930 $\pm$ 95	SMU-68	<i>Melanoides</i> shell	~95	32
	Silts	5,940 $\pm$ 70	Q-1359	Mixed shell	82	
	Coquina	6,135 $\pm$ 115	SUA-503	<i>Bellamya</i> shell	64‡	30
	Coquina	6,420 $\pm$ 125	SUA-499	<i>Bellamya</i> shell	85‡	30
	Beach sands	6,860 $\pm$ 125	Q-1101	Shell	93	11
	Shore terrace	8,240 $\pm$ 295†	SMU-66	<i>Corbicula</i> shell	~95	32
	Coquina	8,655 $\pm$ 175	Q-1114	<i>Melanoides</i> shell	88	11
	Fine sand/silt	9,270 $\pm$ 85	Q-1358	Mixed shell	75‡	
	Shore terrace	9,450 $\pm$ 105	SMU-63	<i>Corbicula</i> shell	~101	32
Ziway-Shala V	Profundal mud	9,550 $\pm$ 170	Gif-3967	Organic fraction (core)	22‡	12–14
	Coquina	9,580 $\pm$ 195	GaK-3386	<i>Melanoides</i> shell	84‡	11
	Shore terrace	9,690 $\pm$ 105	SMU-64	<i>Melanoides</i> shell	~101	32
	Coquina	9,720 $\pm$ 215	Not publ.	Shell	88‡	31
	Profundal mud	9,810 $\pm$ 170	Gif-3966	Organic fraction (core)	23‡	12–14
	Coquina	9,840 $\pm$ 140	SUA-501	<i>Melanoides</i> shell	75‡	
	Coquina	10,350 $\pm$ 150†	SUA-502	<i>Melanoides</i> shell	82‡	
	Marl	12,560 $\pm$ 210†	Gif-3985	Carbonate fraction	57‡	
Ziway-Shala IV						

All ^{14}C dates have been corrected for fractionation using ^{13}C measurements if available or group mean values for similar materials. For locations of dated samples see ref. 12.

* Rich shell concentrate from which fines have been winnowed.

† Stratigraphically inconsistent by $>2\sigma$.

‡ Minimum elevation for corresponding lake level.

the time of overflow. Around 5,000 yr BP, a major recession occurred. Since then, dry conditions have prevailed. A well preserved shoreline at +42 m SD, overlying a charcoal sample dated 2,510 $\pm$ 100 yr BP (Table 1), represents the highest level reached by the three southern lakes during the past few millennia.

The terminal Pleistocene to mid-Holocene period of high lake levels in tropical Africa, culminating around 9,000 yr BP when

precipitation is estimated to have been at least 25% greater than today^{16–18}, is frequently used as an analogue for the future impact of a CO_2 -induced climatic warming^{19,20}. This wet phase has been attributed to the early post-glacial maximum of summer radiation in the Northern Hemisphere^{18,21}, which must have been reinforced by the rise in atmospheric CO_2 content after 13,000 yr BP (refs 22, 23). These two mechanisms, however, operate on too long a time scale to explain the episodes of severe drought, lasting 10^2 – 10^3 yr, described above. Although the relative impact of these events varied from area to area, the data presented here suggest conditions as dry as, or even drier than, today. It is therefore necessary to postulate an additional forcing mechanism, which may be capable of overriding the effects of any future climatic change due to CO_2 . The mechanisms most likely to operate on this time scale include: episodic volcanism, solar variability, cryospheric instability (such as glacial 'surges') and perturbations in sea-surface temperature or salinity^{22–28}.

We thank G. Delibrias and R. L. Otlet for additional ^{14}C dates, and N. J. Shackleton for ^{13}C determinations. Permission to work in Ethiopia was granted by the Institute of Ethiopian Studies, Addis Ababa University, and by the Ethiopian Ministry of Mines. We acknowledge support from the NERC, the Royal Society, CNRS and the Department of Geography, Cambridge University.

Received 3 July; accepted 6 October 1983.

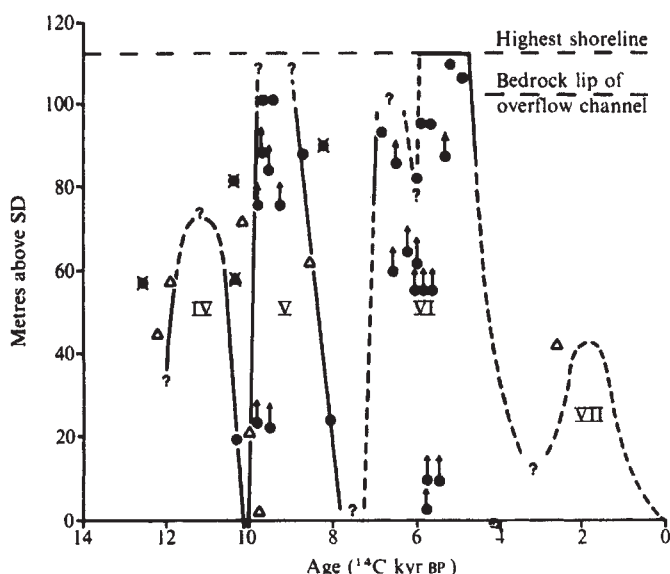


Fig. 2 Water-level curve for palaeolake Ziway-Shala, 14,000–0 yr BP, based on all the available stratigraphical, palaeoecological and ^{14}C data (refs 12–14 and this paper). ●, ^{14}C dates on lacustrine materials; △, ^{14}C dates on nonlacustrine materials; ×, date stratigraphically inconsistent by $>2\sigma$; †, minimum elevation for corresponding lake level; dashed line denotes uncertainty.

- Grove, A. T. & Goudie, A. S. *Nature* **234**, 403–405 (1971).
- Butzer, K. W., Isaac, G. L., Richardson, J. L. & Washbourn-Kamau, C. *Science* **175**, 1069–1076 (1972).
- Servant, M. & Servant, S. *Rev. Géogr. phys. Géol. dyn.* **12**, 63–76 (1970).
- Michel, P. *Mém. Office de la Recherche Scientifique et Technique Outre-Mer* No. 63 (1973).
- Gasse, F. thesis, Univ. Paris-VI (1975).
- Rognon, P. *Rev. Géogr. phys. Géol. dyn.* **18**, 251–282 (1976).
- Butzer, K. W. *Phys. Geogr.* **1**, 42–58 (1980).
- Talbot, M. R. & Delibrias, G. *Earth planet. Sci. Lett.* **47**, 336–344 (1980).
- Haynes, C. V. *Geol. Soc. Am. Abstr. Progr.* **14**, 511 (1982).
- Owen, R. B., Barthelme, J. W., Renaut, R. W. & Vincens, A. *Nature* **298**, 523–529 (1982).
- Grove, A. T., Street, F. A. & Goudie, A. S. *Geogr. J.* **141**, 177–202 (1975).
- Street, F. A. thesis, Univ. Cambridge (1979).

13. Lézine, A.-M. & Bonnefille, R. *Pollen Spores* **24**, 463–480 (1982).
14. Descourtieux-Coquegniot, C. thesis, Univ. Paris-VI (1979).
15. Hamilton, A. *Abstr. X INQUA Congr., Birmingham*, 193 (1977).
16. Street, F. A. *Palaeoecol. Afr.* **11**, 135–143 (1979).
17. Kutzbach, J. *Quat. Res.* **14**, 210–223 (1980).
18. Kutzbach, J. in *Variations in the Global Water Budget* (eds Street-Perrott, F. A., Beran, M. & Ratcliffe, R. A. S.) 371–389 (Reidel, Dordrecht, 1983).
19. Kellogg, W. W. & Schwere, R. *Climate Change and Society* (Westview, Boulder, 1981).
20. Butzer, K. W. *Prof. Geogr.* **32**, 269–278 (1980).
21. Kutzbach, J. *Science* **214**, 59–61 (1981).
22. Delmas, R. J., Ascencio, J.-M. & Legrand, M. *Nature* **284**, 155–157 (1980).
23. Neff, A., Oeschger, H., Schwander, J., Stauffer, B. & Zumbund, R. *Nature* **295**, 220–222 (1982).
24. Eddy, J. A. *Climat. Change* **1**, 173–190 (1977).
25. Flohn, H. *Quat. Res.* **12**, 135–149 (1979).
26. Flohn, H. & Nicholson, S. *Palaeoecol. Afr.* **12**, 3–21 (1980).
27. Flohn, H. in *Variations in the Global Water Budget* (eds Street-Perrott, F. A., Beran, M. & Ratcliffe, R. A. S.) 403–418 (Reidel, Dordrecht, 1983).
28. Street-Perrott, F. A. & Roberts, N. in *Variations in the Global Water Budget* (eds Street-Perrott, F. A., Beran, M. & Ratcliffe, R. A. S.) 331–345 (Reidel, Dordrecht, 1983).
29. Street, F. A. *Proc. VIII Panaf. Congr. Prehist. Quat. Stud.* 143–146 (International Louis Leakey Memorial Institute for African Prehistory, Nairobi, 1980).
30. Gasse, F. & Street, F. A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **24**, 279–325 (1978).
31. Gèze, F. *Bull. géophys. Obs. Addis Ababa* **15**, 119–124 (1975).
32. Haynes, V. & Haas, H. *Radiocarbon* **16**, 368–380 (1974).

Evidence from earliest known erinaceomorph basicranium that insectivorans and primates are not closely related

Michael J. Novacek*, Malcolm C. McKenna*, Nancy A. Neff† & Richard L. Cifelli*

* Department of Vertebrate Paleontology, American Museum of Natural History, New York, New York 10024, USA

† Museum of Natural History, U-43, The University of Connecticut, Storrs, Connecticut 06268, USA

The erinaceomorph insectivorans, which include the living hedgehog and a diversity of archaic taxa^{1–3}, have long been considered to be close relatives or direct ancestors of primates and a variety of other placental mammal orders^{4–6}. We report here on the oldest known erinaceomorph basicranium which provides new evidence against the view that primates and erinaceomorphs share a close common ancestry.

The basicranium is part of a dorsoventrally crushed skull and upper dentition of *Diacodon alticuspsis*, an erinaceomorph from the Lower Eocene (Wasatchian North American Land Mammal Age, about 52 Myr BP) of northern New Mexico⁷. The specimen was collected in 1946 by an American Museum of Natural History field party^{8,9}, but only its dentition (Fig. 1) has been described previously⁷. Although dentally conservative, *Diacodon* is clearly allied with other early Tertiary erinaceomorphs⁷. Skulls of more specialized erinaceomorphs are known from middle Eocene and later records of Europe and North America^{10–12}.

A basicranial feature emphasized as evidence for close primate–erinaceomorph relationship is the incorporation of the petrosal bone in the auditory bulla⁶, a rare condition in eutherians. In those primates where ontogenetic information is available, the bulla is almost entirely formed from the petrosal, although there are small, variable contributions from the ectotympanic¹³. In erinaceomorphs, the bulla typically comprises a petrosal wing and a basisphenoid wing^{12,13}. It has thus been proposed that the common ancestor of primates and erinaceomorphs had a petrosal bullar component⁶. The basicranium of *Diacodon* contradicts this hypothesis: there is no evidence of any petrosal contribution to an auditory bulla. If a bulla was originally present in *Diacodon*, but lost through post-mortem damage¹⁴, it was formed by an independent element (such as an entotympanic) rather than from the petrosal or basisphenoid (Fig. 2b). The petrosal crest (TC in Fig. 2b) in *Diacodon* shows no trace of a broken edge that might be interpretable as the base of a more extensive bullar wing. Similarly, the basisphenoid lacks ridges, inflated areas or

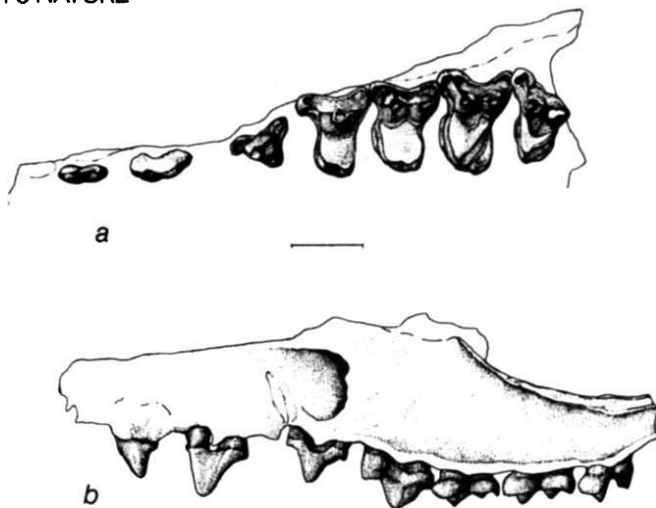


Fig. 1 Upper dentition of *Diacodon alticuspsis*. Occlusal (a) and lateral (b) views of American Museum of Natural History (AMNH) 48597, a left maxillary fragment with P¹–M³ associated with the right maxilla and partial skull (see Fig. 2). P¹ is reconstructed from right side of the same specimen. *D. alticuspsis* was for many years incorrectly referred to the Leptictidae, an extinct group of insectivoran-like forms from North America and Europe. A detailed description of this dentition, a justification for the reference of *Diacodon* to the Erinaceomorpha and a definition of this suborder are provided elsewhere⁷. Scale bars, 3 mm.

broken edges that would suggest bullar development from this element.

It has been suggested that the absence of a bulla, as in soricoid insectivorans and various archaic taxa, was primitive for eutherians^{4,13–15}. The basicranial features of *Diacodon* suggest that early erinaceomorphs either lacked a bulla or had a bulla composed of an element other than the petrosal or basisphenoid. Maintaining a close primate–erinaceomorph tie in the face of this evidence requires the divergence of primates from within Erinaceomorpha, rendering the latter group paraphyletic in spite of other strong defining characters demonstrating its monophyly⁷. More plausibly, the petrosal contributions to the bullae in primates and more modern erinaceomorphs were independently acquired.

Another feature of *Diacodon* relevant to the problem of erinaceomorph–primate affinities is the inferred condition of the carotid circulation. Insectivorans, most primates, some primitive carnivorans, primitive ungulates and microchiropterans are distinguished from most other eutherians and all non-eutherians in having a promontory branch of the internal carotid artery^{15,16}. This vessel is usually conveyed by a well developed groove or tube on the ventral surface of the promontorium. In *Diacodon*, there is no such groove (Fig. 2b), and thus no evidence for a promontory artery. Osseous tracings are not strictly reliable as clues to carotid circulation¹⁷. Nonetheless, an osseous groove or tube carrying a promontory artery is consistently present in living erinaceomorphs. In *Diacodon*, a trough ventral to the abutment between the promontorium and basisphenoid may have carried an entocarotid artery in a medial rather than promontory position (Fig. 2a, b). This pattern is atypical for euprimates and insectivorans, but has recently been reconstructed for the plesiadapiform primate *Ignacius*¹⁸. The promontory branch may merely represent the lateral shift of the 'medial' entocarotid from its primitive position medial to the tympanic cavity¹⁹. If such is the case, early primates and erinaceomorphs retained the primitive carotid condition.

Most proposals for close relatives or ancestors of primates focus on one of the following candidates: (1) specifically erinaceomorph insectivorans^{4–6}; (2) lipotyphlous insectivorans²⁰ (the restricted order including erinaceomorphs, shrews, moles, tenrecs, golden moles, solenodons and several

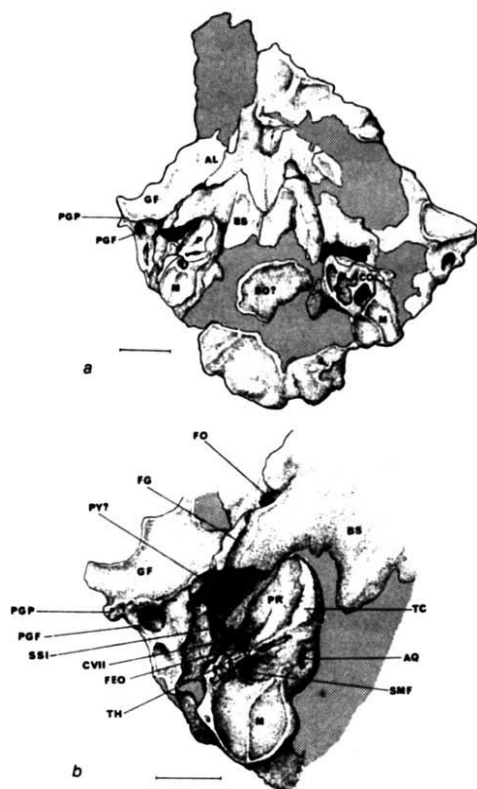


Fig. 2 Ventral views (from slightly different angles) and *D. alticus*, AMNH 48597, posterior region of partial skull (a) and detail of right auditory region (b). Scale bar in a, 5 mm; scale bar in b, 3 mm. Key: AL, alisphenoid; AQ, aqueduct cochleae; BO?, possible basioccipital; BS, basisphenoid; CO, cochlea (damaged on left side); CVII, canal for the facial (VII) nerve; FEO, fenestra ovalis; FG, Glaserian fissure; FO, foramen ovale; GF, glenoid fossa; L, cochlear labyrinth; M, ventral mastoid region of the petrosal bone; P, petrosal; PGP, postglenoid foramen; PGP, postglenoid process; PR, promontorium of the petrosal; PY?, possible piriform fenestra; SMF, stylomastoid foramen; SSI, sulcus for the inferior branch of the stapedia artery; TC, tympanic crest of the petrosal; TH, tympanohyal. Uniform stippling indicates damaged areas. Note that neither BS nor P has prominent tympanic wings that might serve as bullar components. The small tympanic crest (TC) is strikingly similar to the condition found in *Solenodon* and certain soricoids (forms that lack an osseous bulla) and probably represents a primitive character for lipotyphlous insectivores. In other erinaceomorphs, TC is supplanted by a well developed tympanic process of the petrosal, which contributes to the ossified bulla. Although the sulcus for the stapedia artery is present on the surface of the promontorium, there is no indication of a sulcus for the promontory artery, which normally runs from the posteromedial section to the anterolateral apex of the promontorium. The medial surface of the promontorium shows a depression that suggests the pathway of a vessel. Perhaps this served as the sulcus for the medial entocarotid, which may have exited the auditory region through the gap (a middle lacerate foramen?) between PR and BS (a, b). The lack of a promontory groove does not definitely rule out the presence of a promontory artery; in *Solenodon*, this groove is comparatively faint, but the promontory artery follows its usual course.

fossil taxa); (3) tupaiids^{21,22}, or (4) one of several archaic taxa (for example, leptictids or apatemyids) loosely termed 'insectivores'²³. Efforts to choose among these alternatives are undercut by the lack of an inclusive definition for primates. The widely acknowledged primate petrosal bulla and ossified carotid tubing^{6,24} are not present in all fossil forms regarded as primates on the basis of their dental morphology. In *Ignaciuss*, for example, the bony composition of the bulla cannot be identified, and the carotid circulation departs from the general plan in euprimates¹⁸. It has been suggested therefore that plesiadapiforms, including the problematic microsyopsids, be

recognized as an early, but polyphyletic, primate grade¹⁸. However, such a grade concept merely skirts the critical issue of whether a group Primates can be defined at a broader level than that restricted to euprimates (which are defined by the presence of a petrosal bulla). If the dental evidence is actually compelling enough to link microsyopsids and other putative plesiadapiforms with euprimates, a more explicit definition presenting such evidence is surely warranted. Such a grouping would corroborate our thesis that a petrosal bullar element was independently derived within erinaceomorphs and primates, and is therefore not valid evidence for a close relationship between these two taxa.

We thank Z. Kielan-Jaworowska, N. Shubin and A. Wyss for criticisms. Research support was provided by the Frick Laboratory Endowment Fund and the NSF (SPI 79-14876, to M.J.N.).

Received 20 June; accepted 4 October 1983.

- Butler, P. M. *Proc. zool. Soc. Lond.* **118**, 446-500 (1948).
- Krishtalka, L. *Carnegie Mus. nat. Hist. Bull.* **1**, 1-40 (1976).
- Russell, D. E., Louis, P. & Savage, D. E. *Bull. Mus. nat. Hist.* **327**, 129-193 (1975).
- Gregory W. K. *Bull. Am. nat. Hist.* **27**, 1-524 (1910).
- Simpson, G. G. *Bull. Am. Mus. nat. Hist.* **85**, 1-350 (1945).
- Szalay, F. S. in *Phylogeny of the Primates, a Multidisciplinary Approach* (eds Luckett, W. P. & Szalay, F. S.) 91-125 (Plenum, New York, 1975).
- Novacek, M. J. *Contr. Geol. Univ. Wyoming* **20**, 135-149 (1982).
- Simpson, G. G. *Am. J. Science* **246**, 257-282 (1948).
- Simpson, G. G. *Am. J. Science* **246**, 363-385 (1948).
- Maier, W. *Palaeont. Z.* **53**, 38-62 (1979).
- Dechaseaux, C. *Ann. Paléont. (Vertébrés)* **1964**, 83-100 (1964).
- Rich, T. H. V. *Bull. Am. Mus. nat. Hist.* **171**, 1-116 (1981).
- MacPhee, R. D. E. *Contr. Primatol.* **18**, 1-282 (1981).
- Novacek, M. J. *Mammal Rev.* **7**, 131-149 (1977).
- Archibald, J. D. *J. hum. Evol.* **6**, 609-622 (1977).
- Cifelli, R. L. *J. Paleont.* **56**, 795-805 (1982).
- Conroy, G. C. & Wible, J. R. *Folia Primatol.* **29**, 81-85 (1978).
- MacPhee, R. D. E., Cartmill, M. & Gingerich, P. D. *Nature* **301**, 509-511 (1983).
- Presley, R. *Acta anat.* **103**, 238-244 (1979).
- Novacek, M. J. in *Macromolecular Sequences in Systematic and Evolutionary Biology* (ed. Goodman, M.) 3-41 (Plenum, New York, 1982).
- Le Gros Clark, W. E. *The Antecedents of Man* 1st edn (Edinburgh University Press, 1959).
- Szalay, F. S. & Drawhorn, G. in *Comparative Biology and Evolutionary Relationships of Tree Shrews* (ed. Luckett, W. P.) 133-169 (Plenum, New York, 1980).
- McKenna, M. C. *Folia primatol.* **4**, 1-25 (1966).
- Szalay, F. S. & Delson, E. *Evolutionary History of the Primates* (Academic, New York, 1979).

Evidence that disinhibition of brain stem neurones contributes to morphine analgesia

H. L. Fields, H. Vanegas*, I. D. Hentall & G. Zorman

Departments of Physiology and Neurosurgery, University of California, San Francisco, California 94143, USA

Analgesia results when opiates are microinjected into the rostral ventromedial medulla (RVM)¹⁻³. This region, which includes the nucleus raphe magnus and the adjacent reticular formation, is rich in immunoreactive enkephalin-containing neurones and terminals⁴, and contains neurones that project to the spinal cord dorsal horn where they inhibit identified nociceptive spinothalamic tract neurones⁵⁻⁷. Although opiates have previously been reported either to excite or inhibit RVM cells⁸, the possibility of an opiate effect being consistent within a physiologically defined subclass has not been examined. Recently we described a class of neurone in the RVM (the off-cell) that abruptly pauses just before a heat-evoked tail-flick reflex⁹. If off-cells are made to fire continuously by direct electrical stimulation of the RVM, the tail-flick reflex does not occur. We report here that analgesic doses of morphine completely eliminate the pause in firing that precedes the tail-flick reflex. We propose that this disinhibition of off-cells in the RVM is a primary process contributing to opiate inhibition of nociceptor-induced reflexes.

Rats (300-350 g) were initially anaesthetized with sodium pentobarbital (60 mg per kg), intraperitoneally (i.p.), a

* Permanent address: Instituto Venezolano de Investigaciones Científicas, Apartado 1827 Caracas 101, Venezuela.

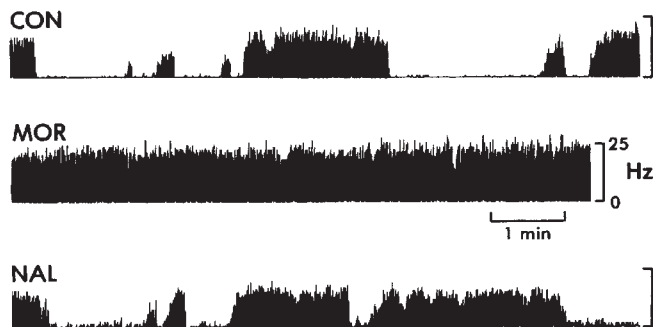


Fig. 1 Effect of 5 mg per kg morphine sulphate i.v. on the resting discharge of an off-cell in the rostral ventromedial medulla. CON, periodic discharge before morphine injection. MOR, 10 min after morphine injection the cell has achieved a steady firing level of ~20 Hz. NAL, 5 min following naloxone administration (1 mg per kg, i.v.), cell discharge has returned to a periodic pattern.

substance which does not affect morphine suppression of nociceptor-induced reflexes¹⁰. A platinum-plated, monopolar steel microelectrode was placed in the dorsal RVM and used for both stimulation and recording. Controlled heat was applied to a small portion of the blackened ventral surface of the tail, and a strain-gauge transducer provided a voltage output when the tail moved. After the initial anaesthesia had lightened, the tail-flick reflex (TF) occurred regularly at tail temperatures above 45 °C. An anaesthetic level which prevented signs of discomfort (for example, orienting or vocalization) but permitted TF was induced by slow infusion of the short-acting barbiturate, sodium methohexital (15–30 mg per kg per h, intravenously (i.v.)). Appropriate recording sites were then found by stimulat-

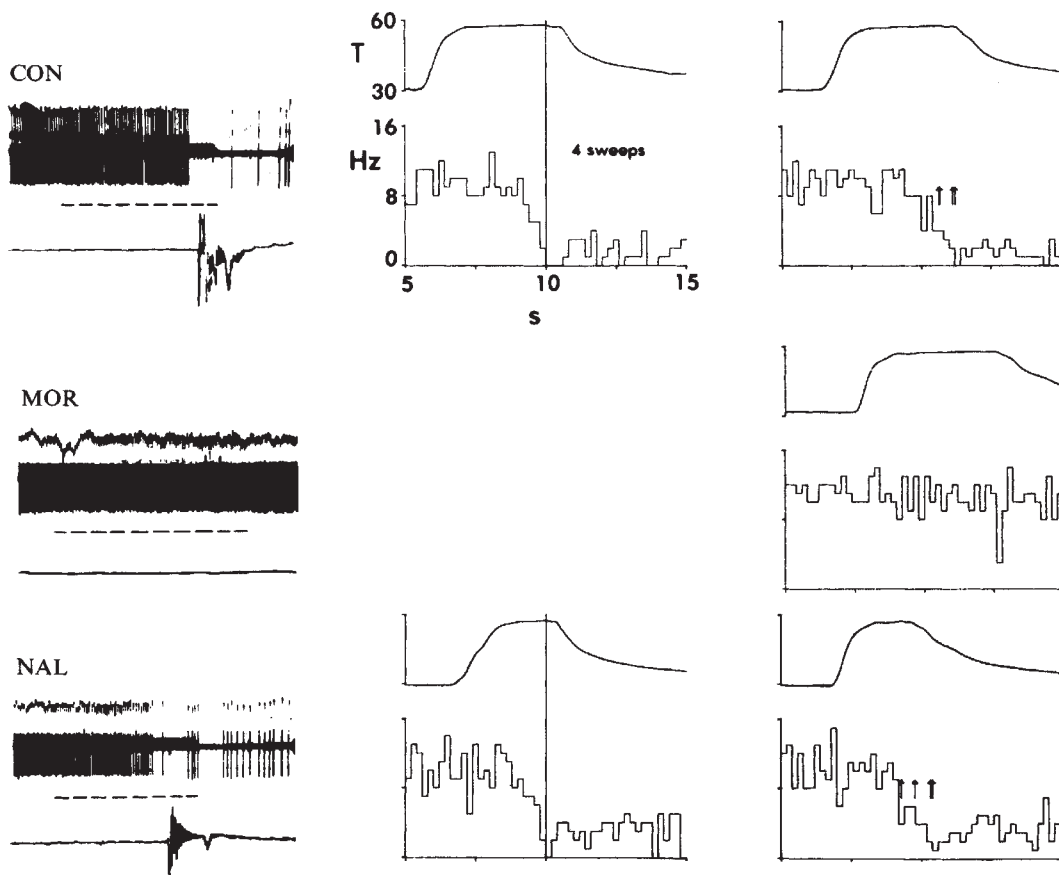
ing (400 μ s pulses, 50 Hz continuous) through the recording microelectrode at successively more ventral sites in the RVM, until the TF was inhibited by stimulus currents of $\leq 10 \mu$ A. Ventral to this point, single unit action potentials were isolated, then stored together with the tail temperature and time of TF occurrence during several repetitions of the heat stimulus. Data were then analysed and displayed as spike histograms plotted with respect to the stimulus (heat) or the response (TF) (Fig. 2).

After identification of an off-cell, its spontaneous discharges were plotted on a chart recorder for about 30 min. Next, morphine sulphate, (5 mg per kg, i.v.) was injected, and the resultant change in resting activity recorded. Typically, after a few minutes, morphine abolished the TF, and responses to noxious heat were then studied again. Finally, the procedure was repeated after naloxone (1 mg per kg, i.v. or i.p.) had restored the TF. Rats were killed with an overdose of barbiturate. All recording sites were verified histologically in Neutral Red stained sections.

Complete physiological and pharmacological data were obtained from nine cells characterized as off-cells. Only one cell was studied in each rat. Eight cells were within the boundaries of the nucleus raphe magnus and one was more lateral, in the reticular formation. All off-cells in the nucleus raphe magnus had an irregular periodic pattern of spontaneous discharge before morphine administration (Fig. 1). Peak discharge rates ranged over 10–18 Hz. Following morphine, cell activity became aperiodic, with average firing rates slightly higher than the peak level before morphine administration. The mean resting discharge of the one cell located outside the nucleus raphe magnus was decreased by morphine.

In all nine cells, morphine completely blocked inhibition of the off-cell while eliminating the TF elicited by heat (Fig. 2).

Fig. 2 Effect of morphine on off-cell discharge during the tail-flick reflex (TF) induced by noxious heat. CON, MOR, NAL lines as in Fig. 1. Left column, single oscilloscope sweeps, for CON and NAL periods the pause in unit activity (upper trace) and the occurrence of the tail-flick (lower trace) are illustrated. The MOR trace illustrates the blockade of the pause and TF by morphine sulphate, 5 mg per kg. The interrupted line immediately below the spike trace indicates the time of application of heat to the tail. Full sweep is 10 s. Average temperature (T, top line) and mean spike frequency (Hz, bottom, 200 ms bins) are plotted in the centre column with respect to tail flick time (vertical line) and, in the right column, with respect to the time at which tail temperature reached 45 °C (small vertical arrows indicate the time of tail-flick). This figure illustrates that there is a sharper correlation of the pause onset with TF occurrence than with the temperature rise. Since morphine abolished the TF, the absence of the pause cannot be illustrated by plotting spike data with reference to TF. These data, averaged over four heat trials, confirm the single sweep data showing that the flick and cell pause are abolished by morphine (MOR) and restored by naloxone (NAL).



Naloxone restored the TF and the off-cell pause in all nine cells.

Nicoll and colleagues have observed that opiates inhibit local inhibitory neurones at several central nervous system sites¹¹. They and others¹² have proposed that the resultant disinhibition of projection neurones is a common function of central nervous encephalineric synapses. Morphine could produce analgesia by a similar disinhibition of off-cells, some of which we recently showed to project to the spinal cord¹³.

The identity of the neural elements which inhibit the off-cell before the TF is not known. As we gave morphine systemically, it is not possible to specify where its initial action occurs. Furthermore, there is evidence that morphine, when given systemically, also has actions at sites outside the RVM (for example, the midbrain and spinal cord) which contribute to its analgesic effect^{5,6}. However, naloxone microinjected into the RVM reverses the analgesic effect of systemic morphine², indicating that a significant part of the analgesic action of systemic opiates must be exerted directly on neuronal structures located within the RVM.

Off-cells in the RVM have the following characteristics which indicate that they inhibit nociceptor-driven neurones in the spinal cord: (1) they are driven by electrical stimulation of RVM near threshold for analgesia¹⁴; (2) they turn off just prior to tail-flick; (3) as we show here, their off-response disappears and reappears in parallel with the TF when opiate agonists and antagonists are administered; (4) they project to the spinal cord¹³; and (5) with stimulation of the midbrain periaqueductal grey, the threshold for excitation of off-cells and suppression of the TF is identical (unpublished studies in our laboratory). Thus, in all situations studied, tail-flick only occurs when off-cells pause. We conclude that off-cells are critical elements in pain modulation and that their disinhibition contributes to the analgesic action of systemically administered opiates.

This work was supported by PHS grant DA01949. We thank John Bry, Nicholas Barbaro, Stephen Guinn and Allan Basbaum.

Received 9 August; accepted 20 September 1983.

1. Akaike, A., Shibata, T., Satoh, M. & Takagi, H. *Neuropharmacology* **17**, 775-778 (1978).
2. Azami, J., Llewellyn, M. D. & Roberts, M. H. T. *Pain* **12**, 229-246 (1982).
3. Dickenson, A. H., Oliveras, J. L. & Besson, J. M. *Brain Res.* **170**, 95-111 (1979).
4. Hökfelt, T., Ljungdahl, A., Terenius, L., Elde, R. & Nilsson, G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3081-3085 (1977).
5. Basbaum, A. I. & Fields, H. L. *Ann. Neurol.* **4**, 451-462 (1978).
6. Yaksh, T. L. & Rudy, T. A. *Pain* **4**, 299-359 (1978).
7. Willis, W. D., Haber, L. H. & Martin, R. F. *J. Neurophysiol.* **40**, 968-981 (1977).
8. Gebhart, G. F. *Pain* **12**, 93-140 (1982).
9. Fields, H. L., Hentall, I. D. & Zorman, G. *J. Neurosci. Abstr.* **8**, 806 (1982).
10. Geller, E. B., Durlflosky, L., Cowan, A., Harakal, C. & Adler, M. W. *Life Sci.* **25**, 139-146 (1979).
11. Nicoll, R. A., Alger, B. E. & Jahr, C. E. *Nature* **287**, 22-25 (1980).
12. Frederickson, R. C. A. & Geary, L. E. *Prog. Neurobiol.* **19**, 19-69 (1982).
13. Vanegas, H., Barbaro, N. M. & Fields, H. L. *Soc. Neurosci. Abstr.* **9**, 3 (1983).
14. Hentall, I. D., Zorman, G. & Fields, H. L. *Soc. Neurosci. Abstr.* **7**, 734 (1981).

Evidence for a central component of post-injury pain hypersensitivity

Clifford J. Woolf

Department of Anatomy, University College London, London WC1E 6BT, UK

Noxious skin stimuli which are sufficiently intense to produce tissue injury, characteristically generate prolonged post-stimulus sensory disturbances that include continuing pain, an increased sensitivity to noxious stimuli and pain following innocuous stimuli. This could result from either a reduction in the thresholds of skin nociceptors (sensitization)^{1,2} or an increase in the excitability of the central nervous system so that normal inputs now evoke exaggerated responses^{3,4}. Because sensitization of peripheral receptors occurs following injury⁵⁻⁷, a peripheral mechanism is widely held to be responsible for

post-injury hypersensitivity. To investigate this I have now developed an animal model where changes occur in the threshold and responsiveness of the flexor reflex following peripheral injury that are analogous to the sensory changes found in man. Electrophysiological analysis of the injury-induced increase in excitability of the flexion reflex shows that it in part arises from changes in the activity of the spinal cord. The long-term consequences of noxious stimuli result, therefore, from central as well as from peripheral changes.

To investigate, in the unanaesthetized rat, the effects of an injury that would be expected to produce pain hypersensitivity in man, it has been necessary to perform the experiments in chronically decerebrate animals. Decerebration was performed under pentobarbital anaesthesia in eight rats. With body temperature control and orogastric tube feeding the decerebrate rats survived for several weeks⁸. Within 24 h of the decerebration the animals exhibited intact spinal and brain-stem reflexes. These included flexor withdrawal reflexes, vocalization and orientation to the site of the stimulus, following noxious stimulation of the hindpaw. The mechanical and thermal thresholds of the flexor reflex remained stable until a localized thermal injury to the lateral edge of the foot was produced. Figure 1 illustrates the fall in the mechanical threshold and thermal response latency resulting from the injury. The injury also increased the responsiveness of the reflex, from a brief flick (~1 s) following a suprathreshold stimulus pre-injury, to a sustained flexion (~40 s) post-injury. Peripheral tissue injury in the decerebrate rat produces, therefore, changes in the threshold and responsiveness of the flexion reflex that parallel the sensory disturbances found in man.

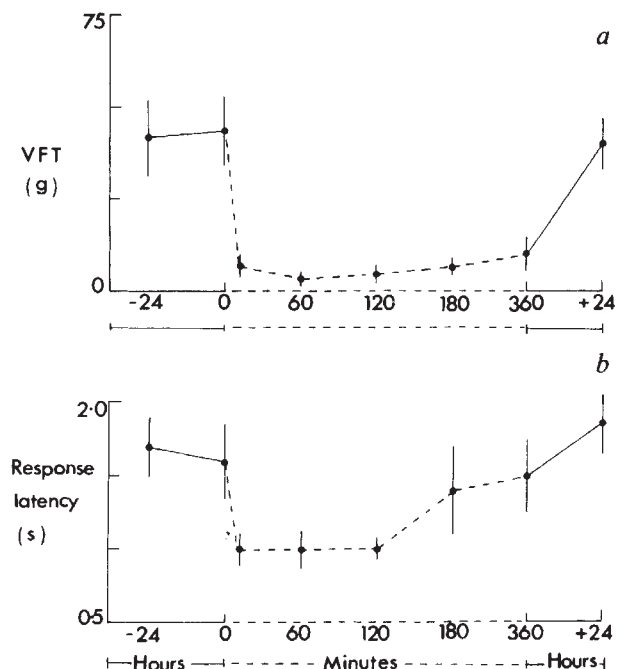


Fig. 1 Alterations in the threshold and responsiveness of the hindlimb flexion withdrawal reflex in eight chronic decerebrate rats produced by a thermal injury (radiant heat to lateral edge of foot, 75 °C for 60 s) at time 0. *a* Shows the reduction in the mechanical threshold, tested with Von Frey hairs (VFT) and expressed in grams, for eliciting flexion following stimulation of the dorsum of the foot adjacent to the site of injury, and its recovery to pre-injury levels within 24 h. *b* Shows the latency of response following immersion of the rats left hindpaw in water at 49 °C before and after that foot was subjected to a thermal injury. All results are expressed as means $\pm$ s.e.m. X-axis: solid, hours; dotted, minutes. The baseline mechanical thresholds of the decerebrate rats were lower than in intact animals but above the thresholds of high-threshold cutaneous mechanoreceptors¹⁰. The baseline thermal response latencies in intact and decerebrate rats were identical.

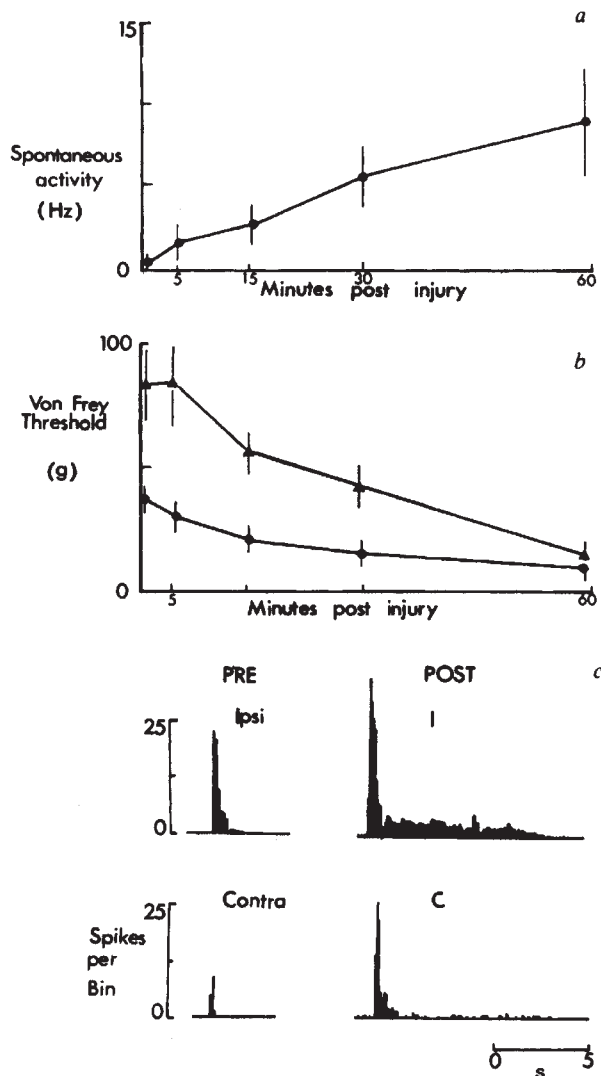


Fig. 2 *a*, Time course of the increase in the spontaneous activity of biceps femoris α motoneurone efferents ($n = 25$) after thermal injury to the ipsilateral foot at time 0. *b*, Mechanical thresholds required to elicit responses from the plantar surface of the ipsilateral ($\bullet$) ($n = 25$) and contralateral ($\blacktriangle$) ($n = 13$) foot determined by testing with Von Frey hairs. The ipsilateral tests were performed adjacent to but not on the site of the injury. *c*, Examples of responses produced in a single biceps femoris efferent by a standard pinch (150 g mm^{-2}) to the plantar surface of the ipsilateral and contralateral second toe, before a thermal injury (PRE) to the lateral side of the foot and 30 min after the injury (POST). Bin width was 500 ms. Note the bilateral increase in response amplitude and duration.

Electrophysiological experiments to examine the post-injury hypersensitivity of the flexion reflex were performed in 28 decerebrate-spinal adult Wistar rats. The animals were decerebrated under Althesin anaesthesia, which was then discontinued and the animals ventilated under gallamine paralysis. Spinalization was performed at T10–T11. The nerve to the biceps femoris muscle and the sural nerve were exposed and covered with mineral oil. The biceps femoris nerve was cut distally, desheathed and small filaments teased out and placed on a recording electrode.

A total of 28 single biceps femoris α -motoneurone efferents were studied. The efferents had a low or absent spontaneous activity and high threshold mechanoreceptive cutaneous receptive fields on the ipsilateral foot which extended in 54% to the tail and contralateral foot⁹. Sixty-eight per cent of the units also fired in response to noxious thermal stimuli ($<5^\circ\text{C}$, $>49^\circ\text{C}$) applied to the ipsilateral foot.

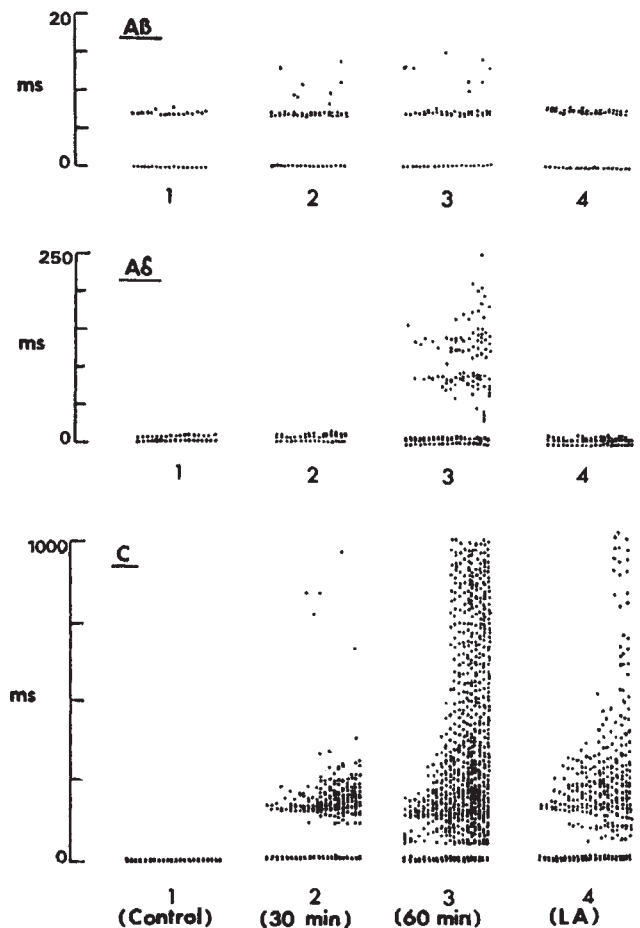


Fig. 3 Raster dot displays of a single biceps femoris unit activated by stimulation of the sural nerve before an ipsilateral thermal injury (control), 30 and 60 min post-injury and 10 min after the injured foot had been completely anaesthetized with Xylocaine. Each dot represents an action potential. The vertical scale is the latency of the response after the sural nerve stimulus. (The stimulus artefact is present at time 0), while the horizontal scale is real time (from left to right) with a new stimulus every 2 s. The sural nerve was stimulated at strengths which only activated Aβ fibres ($100 \mu\text{A}$, $50 \mu\text{s}$), Aβ and Aδ fibres ($250 \mu\text{A}$, $50 \mu\text{s}$) and Aβ, Aδ and C fibres (5 mA , $500 \mu\text{s}$). Note the different timescales used for monitoring the Aβ, Aδ and C evoked responses. In the pre-injury state, only an Aβ input was evoked. Thirty minutes after the injury a C response begins to occur, while at 60 min both Aδ and C evoked bands of activity are present. (Note the progressive response increment or wind-up of the C responses.) Ten minutes following the local anaesthetic (administered 80 min post-injury) the sural C-evoked responses remain higher than before the injury although not as high as immediately before the local anaesthetic.

A thermal injury localized to the lateral edge of the foot, identical to that used in the chronic decerebrate rats, was produced in all 28 animals. In 25 rats the injury was made ipsilateral to the recording site, in three it was contralateral. An immediate area of erythema developed in the skin subjected to the heat stimulus. By 5–10 min oedematous changes in the skin appeared and the swelling and the erythema slowly spread from the site of the stimulus to include most of the foot within 1 h.

The injury resulted in marked changes in the ipsilateral flexor efferents. Their spontaneous activity increased (Fig. 2*a*) and significant falls in the Von Frey hair thresholds for evoking a response from both the ipsi and contralateral mechanoreceptive fields occurred by 1 h (Fig. 2*b*). Associated with these changes was a bilateral increase in the amplitude of response to a standard pinch (150 g mm^{-2}) and its duration (Fig. 2*c*). The afterdischarge to an ipsilateral pinch increased from $2.8 \pm 0.3 \text{ s}$ pre-injury to $14.2 \pm 3 \text{ s}$ post-injury ($n = 25$). Similar changes occurred to the

response evoked by a standard thermal stimulus (50 °C for 10 s). In addition to the changes in existing receptive fields, the size of the cutaneous receptive fields also increased in 72% of the units with an expansion to the ipsilateral hindquarter and the appearance of novel or larger contralateral inputs. One hour post-injury 84% of all units had bilateral receptive fields as opposed to 52% pre-injury. The number of units with contralateral thermoreceptive fields increased from 4% to 32%. Contralateral thermal injuries produced no changes in the responses of three animals so a generalized systemic effect of the injury is ruled out.

Sensitization of peripheral nociceptors adjacent to the injury^{10,11} is probably responsible for some of the changes in the ipsilateral receptive fields of the flexor efferents, but the greater responses at a lower threshold from the contralateral fields and the expansion of the receptive fields, cannot be accounted for by such a sensitization. To examine further the injury-induced excitability change in the flexor efferents the responses to sural nerve stimulation were investigated. Thirty minutes after the injury, 10 of 18 animals tested exhibited a very substantial increase in sural-evoked response. The increased responses were most prominent to A δ and C volleys. In five animals A δ and C stimulation even began to evoke responses post-injury that had not been present before the injury (Fig. 3).

The increased responsiveness of the flexor reflex to natural and electrical inputs post-injury may be either the result of a summation of a continuing afferent input from the site of the injury with previously subthreshold inputs or the result of an injury-triggered increase in the excitability of the spinal cord. To test these possibilities the following experiment was performed on six single flexor efferents. The efferents had all exhibited an expansion of their receptive fields to the contralateral foot or tail by 1 h post-injury and in four these was also a marked increase in their sural evoked responses. Local anaesthetic (2% Xylocaine) was then injected into the ipsilateral foot to produce a complete sensory block of the entire foot and ankle region including the site of the injury and all surrounding inflamed tissue. The adequacy of the sensory block was tested by the development within 5 min of a failure to evoke any responses by natural or electrical stimulation. Ten to fifteen minutes after the sensory block the injury-induced contralateral receptive fields remained present in five of the six units and the sural-evoked response remained above pre-injury levels in three of four units (Fig. 3). Peripheral injury does therefore trigger an increase in the excitability of the spinal cord. Although the responsiveness of the efferents during the sensory block remained above pre-injury levels, in two units there was a small decline in the responses (Fig. 3). Either the excitability increase has a half life of about 10 min or an afferent input from the site of the injury contributes to the changes.

If the changes found here in the flexor reflex parallel changes in the sensory input to the brain then pain hypersensitivity following injury may be due to changes within the central nervous system as well as at the site of the injury. Precisely what neuronal mechanisms are responsible for the central changes requires further study but the changes in the responsiveness and functional connectivity of the spinal cord found here mean that its activity is determined by previous as well as by present inputs.

I thank P. D. Wall, S. McMahon and M. Fitzgerald for useful advice. The work was supported by the MRC.

Received 4 May; accepted 14 October 1983.

1. Perl, E. R. in *Advances in Research and Therapy*, Vol. 1, 17–28 (Raven, New York, 1970).
2. Lewis, T. *Pain* (Macmillan, London, 1942).
3. Sturge, W. A. *Brain* **5**, 494–510 (1883).
4. Hardy, J. D., Wolff, H. G. & Goodell, H. J. *clin. Invest.* **29**, 115–146 (1950).
5. Perl, E. R., Kumazawa, T., Lynn, B. & Kenins, P. *Prog. Brain Res.* **43**, 236–276 (1976).
6. Fitzgerald, M., *J. Physiol., Lond.* **297**, 207–216 (1979).
7. LaMotte, R. H., Thalhammer, J. G., Törebjörk, H. & Robinson, C. J. *J. Neurosci.* **2**, 765–781 (1982).
8. Woods, J. W. *J. Neurophysiol.* **27**, 635–641 (1964).
9. Swett, J. & Woolf, C. J. *J. Physiol. Lond.* **341**, 20P–21P (1983).
10. Lynn, B. & Carpenter, S. E. *Brain Res.* **287**, 493–508 (1982).
11. Fleischer, E., Handwerker, H. O. & Joukhadar, S. *Brain Res.* **267**, 81–92 (1983).

Substance P in the ascending cholinergic reticular system

S. R. Vincent, K. Satoh, D. M. Armstrong* & H. C. Fibiger

Division of Neurological Sciences, Department of Psychiatry, The University of British Columbia, 2255 Wesbrook Mall, Vancouver, British Columbia, Canada V6T 1W5

* Department of Pathology, Albert Einstein College of Medicine, Bronx, New York 10461, USA

The neocortex receives a major cholinergic innervation from magnocellular neurones in the basal forebrain^{1–3}. However, an ascending cholinergic reticular system has also been postulated to arise from acetylcholinesterase (AChE)-containing neurones in the midbrain and pontine tegmentum^{4–5}. Lesions of this region decrease both AChE and choline acetyltransferase (ChAT) in various forebrain areas^{6,7}, and recent immunohistochemical studies have identified a group of ChAT-containing cell bodies in the midbrain reticular formation and dorsolateral pontine tegmentum^{8,9}. Here we have combined retrograde tracing with ChAT immunohistochemistry to demonstrate that this tegmental cholinergic cell group also directly innervates the cerebral cortex. Other immunohistochemical studies have indicated that the neuropeptide substance P is also present in certain cells in the laterodorsal tegmentum¹⁰, and these too appear to project to the forebrain^{11,12}. We have therefore performed immunohistochemistry for both ChAT and substance P and have discovered that a subpopulation of the ascending cholinergic reticular neurones contains substance P. Thus, peptide–cholinergic coexistence, previously noted in peripheral neurones, also occurs in the brain.

Indirect immunofluorescence techniques were used to identify the cholinergic and substance P-containing neurones of the midbrain and pontine tegmentum of young adult male rats (150–200 g, Wistar). To increase the levels of ChAT and substance P in cell bodies, the animals were given an injection of the mitotic inhibitor colchicine (80 μ g in 20 μ l saline) into the third ventricle, 48 h before perfusion. The brains were fixed by perfusion through the ascending aorta with 50 ml of saline followed by 250 ml of 4% paraformaldehyde in 0.1 M phosphate-buffered saline. Brains were removed and postfixed for 2 h in the same solution and then rinsed for at least 48 h in buffered 15% sucrose. Coronal sections were cut at 25 μ m thickness on a freezing microtome and collected in 50 mM Tris buffered saline (TBS).

Adjacent sections were processed for immunohistochemistry for ChAT and substance P. Indirect immunofluorescence of substance P was performed with an antibody raised in guinea pig¹³, diluted 1:200 in TBS containing 0.5% Triton X-100 and 0.1% bovine serum albumin (BSA). Sections were incubated for 48 h at 4 °C in the primary antibody, rinsed in TBS and incubated at room temperature for 1 h in affinity-purified rabbit anti-guinea pig IgG labelled with rhodamine (TRITC; Zymed) diluted 1:20, rinsed in TBS and mounted in glycerol containing 5% *n*-propyl gallate, buffered with Tris-Cl to pH 8.

A well-characterized monoclonal antibody to ChAT¹⁴ was used to identify the cholinergic cell bodies using a biotin–avidin immunofluorescence procedure. Sections were incubated at 4 °C for 48 h in ammonium sulphate-fractionated culture supernatant of hybridoma AB8¹⁴ diluted 1:50 in TBS with Triton and BSA as above. After rinsing in TBS for 20 min the sections were incubated for 1 h at room temperature in biotinylated rabbit anti-rat IgG (Vector Laboratories), diluted 1:50, rinsed and then incubated in fluorescein isothiocyanate (FITC)-conjugated avidin (Vector), diluted 1:100 in TBS.

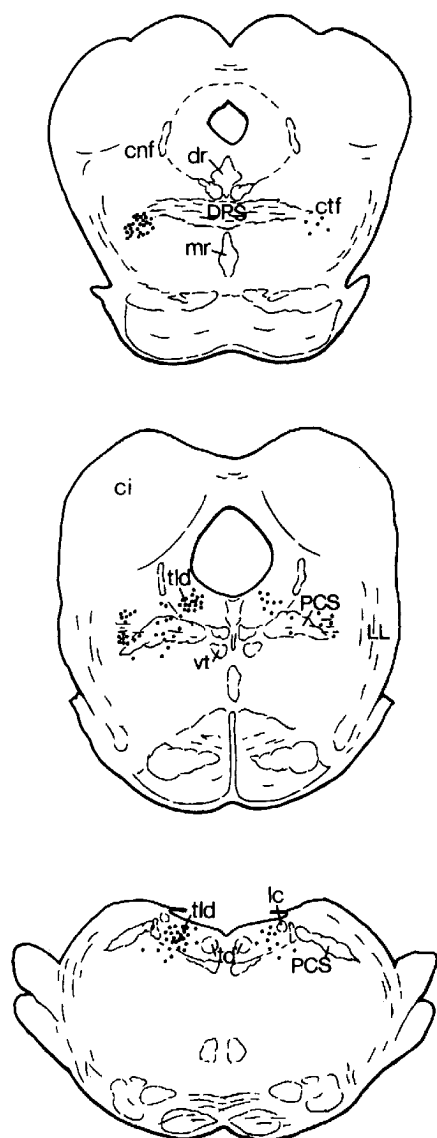


Fig. 1 Schematic frontal sections of the caudal mesencephalon and pons of the rat indicating the distribution of the tegmental cholinergic cell group (black dots on left side) and the neurones containing both acetylcholine and substance P (right side). ci, inferior colliculus; cnf, cuneiform nucleus; ctf, central tegmental field (Berman); dr, dorsal raphe nucleus; lc, locus coeruleus; mr, median raphe nucleus; td, dorsal tegmental nucleus (Gudden); tld, laterodorsal tegmental nucleus (Castaldi, as modified from ref. 15); vt, ventral tegmental nucleus (Gudden); DPS, decussation of the superior cerebellar peduncle; LL, lateral lemniscus; PCS, superior cerebellar peduncle.

Sections were also prepared that were simultaneously stained for both ChAT and substance P. For this procedure sections were incubated in a mixture of the two primary sera at their respective final dilutions. The second and third steps of the ChAT procedure were performed next, followed by the affinity-purified TRITC-labelled anti-guinea pig IgG step. By means of this technique and the use of the appropriate dichroic mirrors and filters on the fluorescence microscope, the red substance P-positive cells and the green ChAT-positive cells could be separately visualized in a single section.

A large group of ChAT-positive neurones was found in a circumscribed region of the midbrain and rostral pons (Fig. 1). Rostrally, this cell group begins in the midbrain reticular formation just caudal to the substantia nigra, and extends dorso-caudally in the lateral portion of the mesencephalic and pontine reticular formation. In the rostral pons, the ChAT neurones

extend medially in the dorsolateral pontine tegmentum and enter the periventricular grey, forming a dense cluster in the laterodorsal tegmental nucleus (TLD) (Castaldi as modified from ref. 15). In adjacent sections, many substance P-positive neurones displaying a morphology similar to that of the ChAT cells were found in all these regions, particularly within the TLD.

When sections were simultaneously stained for both ChAT and substance P, many of the cholinergic neurones displaying green FITC fluorescence also displayed red TRITC fluorescence, indicative of substance P immunoreactivity (Fig. 2). This was not due to cross-reactivity of any of the sera because many ChAT-intense neurones did not display substance P immunoreactivity, and also, the omission of either one of the primary antibodies resulted in only single-labelled cells. Analysis of over 1,000 cholinergic neurones by double staining indicated that at least one-third of the neurones in the midbrain and pontine cholinergic cell column contain substance P immunoreactivity. No substance P-containing neurones were found in these regions that did not also contain ChAT. The frequency of the double-labelled cells varied in different areas, with about 40% of the ChAT cells in the TLD being also substance P positive while in the lateral midbrain reticular formation only about 20% of the cholinergic neurones also contained the peptide.

Neurones in the TLD have recently been shown to project to the medial frontal cortex¹⁶, and some of these neurones appear to contain substance P¹². As a cortical cholinergic innervation arises from neurones in the basal forebrain¹⁻³, we thought it of particular interest to determine if the TLD projection to the frontal cortex also contains ChAT. To study this possibility, we combined a retrograde tracing technique with immunohistochemistry for ChAT and substance P. Three rats received stereotaxic unilateral injections of the retrograde fluorescent tracer True blue¹⁷ into the medial frontal cortex. Four days later the animals received colchicine as described above and were perfused 2 days later. Sections were cut through the injection site and the midbrain and pons and stained for both ChAT and substance P using the FITC and TRITC markers, respectively. The sections were mounted with paraffin oil to preserve the tracer fluorescence and examined and photographed with the appropriate filter systems for the tracer and the two immunostains.

After frontal cortex injections of the tracer, many True blue-labelled neurones were found within the TLD both on the injected side, and to a lesser extent on the contralateral side. In contrast, the retrogradely labelled cells in the rostral locus coeruleus, just lateral to the caudal TLD, were found predominantly only on the injected side. Immunohistochemical examination showed that the retrogradely labelled TLD neurones contained ChAT or both ChAT and substance P immunoreactivities (Fig. 3). Many of the ChAT- and substance P-labelled cells in the lateral midbrain and pontine reticular formation were also retrogradely labelled by the cortical injections.

The present results provide direct morphological evidence for the coexistence of the neuropeptide substance P with acetylcholine in an ascending reticular system. In addition, we have demonstrated that many of the neurones in this cholinergic cell group project to the cerebral cortex. Thus, this system represents a second source of cortical cholinergic innervation in addition to the well described projection originating in the basal forebrain¹⁻³, and may also be a primary source for the substance P fibres in the cerebral cortex^{10,12,18}. The basal forebrain cholinergic neurones are known to degenerate in Alzheimer's disease^{19,20}, and recently loss of cortical substance P in this disorder has also been cited²¹. This raises the interesting possibility that the cholinergic cells of the midbrain and pontine tegmentum, which contain substance P, may also be involved in this disease.

Various studies have indicated that functional interactions between substance P and central cholinergic mechanisms may occur. Acetylcholine and substance P have a similar subcellular distribution in the brain^{22,23}. Substance P has been shown to

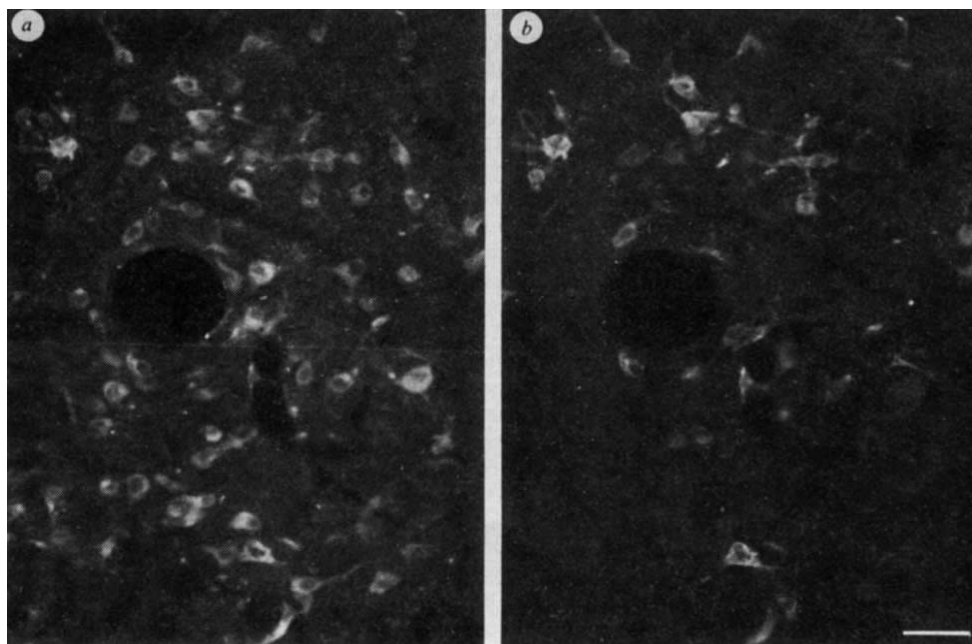


Fig. 2 Immunofluorescence micrographs of a section through the laterodorsal tegmental nucleus showing the cholinergic neurons using a green fluorescing FITC marker for ChAT-positive cells (*a*) and the substance P-immunoreactive neurons, on the same section, identified with a red fluorescing TRITC label (*b*). Note that all of the substance P-positive neurones are also ChAT positive, although many cholinergic neurones do not display substance P immunoreactivity. Scale bar, 50 μ m.

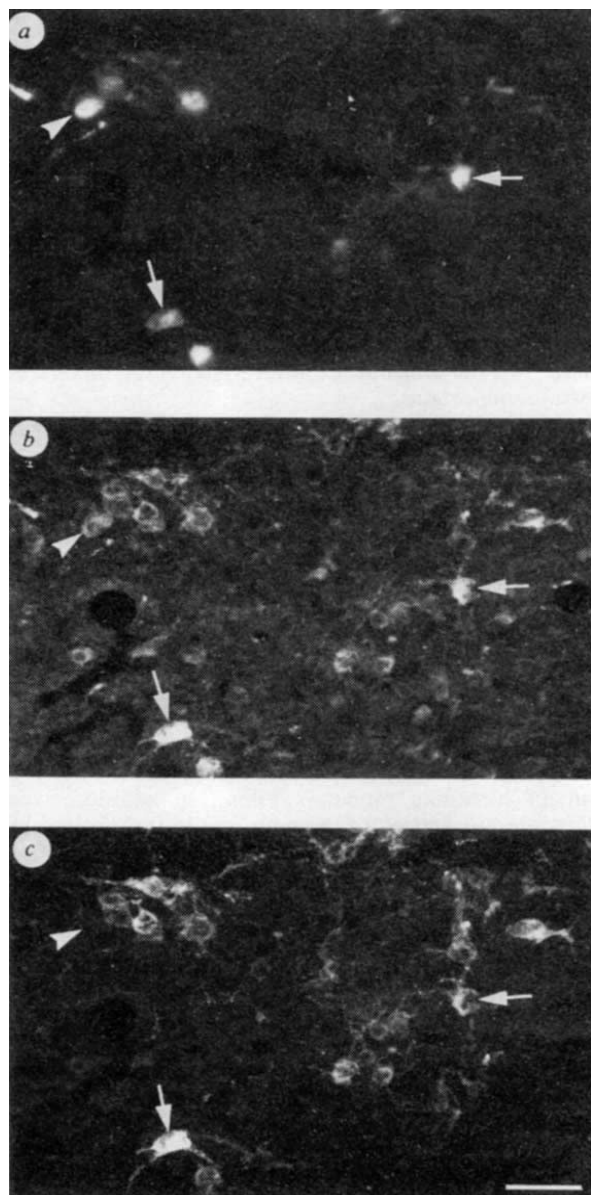


Fig. 3 Fluorescence micrographs of the laterodorsal tegmental nucleus illustrating retrogradely labelled cells (*a*), cholinergic neurones (*b*) and substance P-immunoreactive neurones (*c*) on a single section. *a*, Neurones retrogradely labelled after an injection of the fluorescent tracer True blue into the medial frontal cortex were photographed in the U dichroic mirror, filter illumination system (Olympus) providing excitation light of approximately 334 and 365 nm. *b*, ChAT-positive cells were labelled with FITC and photographed under B system illumination (405, 435 and 490 nm excitation). *c*, Substance P-positive cells were labelled with TRITC under G system illumination at 546 nm. Note that the retrogradely labelled neurones contain both ChAT and substance P (arrows) or only ChAT and not substance P (arrowheads). Scale bar, 50 μ m.

block the nicotinic excitation of Renshaw cells^{24,25} and can be hydrolysed by AChE²⁶. Nicotinic receptors have been found to regulate substance P release in the substantia nigra²⁷, and a cholinergic mechanism seems to mediate the increase in hypothalamic blood flow induced by substance P²⁸. In addition, substance P may regulate hippocampal acetylcholine turnover²⁹. The present demonstration of substance P in some of the tegmental cholinergic neurones projecting to the forebrain may provide an anatomical basis for such interactions.

The presence of both acetylcholine and vasoactive intestinal peptide in certain secretomotor autonomic neurones innervating exocrine glands has proved to be a useful model in which to study the functional significance of transmitter coexistence³⁰. Our discovery of a coexistence of substance P and acetylcholine in the brain may now allow such functional studies to be extended to include central neurones.

This work was supported by grants from the MRC, NIH (NIH-A6/NS 02478) and the McKnight Foundations, Inc. S.R.V. is an MRC scholar.

Received 16 August; accepted 27 September 1983.

1. Divac, I. *Brain Res.* **93**, 385–398 (1975).
2. Lehmann, J., Nagy, J. I., Atmadja, S. & Fibiger, H. C. *Neuroscience* **5**, 1161–1174 (1980).
3. Fibiger, H. C. *Brain Res. Rev.* **4**, 327–388 (1982).
4. Shute, C. C. D. & Lewis, P. R. *Nature* **199**, 1160–1165 (1963).
5. Shute, C. C. D. & Lewis, P. R. *Brain* **90**, 497–522 (1976).
6. Hoover, D. B. & Jacobowitz, D. M. *Brain Res.* **170**, 113–122 (1979).
7. Rotter, A. & Jacobowitz, D. M. *Brain Res. Bull.* **6**, 525–529 (1981).
8. Kimura, H., McGeer, P. L., Peng, J. H. & McGeer, E. G. *J. comp. Neurol.* **200**, 151–201 (1981).
9. Kimura, H. & Maeda, T. *Brain Res. Bull.* **9**, 493–499 (1982).
10. Ljungdahl, Å., Hökfelt, T. & Nilsson, G. *Neuroscience* **3**, 861–943 (1978).
11. Sakanaka, M. *et al. Brain Res.* **230**, 351–355 (1981).
12. Sakanaka, M. *et al. Brain Res.* **259**, 123–126 (1983).

13. Vincent, S. R., Kimura, H. & McGeer, E. G. *J. comp. Neurol.* **199**, 113–123 (1981).
14. Levey, A. I. *et al. J. Neurosci.* **3**, 1–9 (1983).
15. Shimizu, N. *et al. Archs Histol. Japan* **41**, 103–112 (1978).
16. Divac, I. *et al. Neuroscience* **3**, 785–796 (1978).
17. Björklund, A. & Skagerberg, G. *J. Neurosci. Meth.* **1**, 261–277 (1979).
18. Paxinos, G., Emson, P. C. & Cuello, A. C. *Neurosci. Lett.* **7**, 127–131 (1978).
19. Whitehouse, P. J. *et al. Ann. Neurol.* **10**, 122–126 (1981).
20. Nagai, T. *et al. Neurosci. Lett.* **36**, 195–199 (1983).
21. Crystal, H. A. & Davies, P. *J. Neurochem.* **38**, 1781–1784 (1982).
22. Kataoka, J. *Jap. J. Physiol.* **12**, 81–96 (1982).
23. Ryall, R. W. *J. Neurochem.* **11**, 131–145 (1964).
24. Belcher, G. & Ryall, R. W. *J. Physiol., Lond.* **272**, 105–119 (1977).
25. Krnjevic, K. & Lekie, D. *Can. J. Physiol. Pharmac.* **55**, 958–961 (1977).
26. Chubb, I. W., Hodgson, A. J. & White, G. H. *Neuroscience* **5**, 2065–2072 (1980).
27. Torrens, Y. *et al. Eur. J. Pharmac.* **71**, 383–392 (1981).
28. Klugman, K. P. *et al. Br. J. Pharmac.* **71**, 623–629 (1980).
29. Maltbe-Sørensen, D., Cheney, D. L. & Costa, E. *J. Pharmac. exp. Ther.* **206**, 21–28 (1978).
30. Lundberg, J. M. *Acta Physiol. Scand.* **112** (Suppl. 496) 1–57.

β -Adrenoceptor agonists increase membrane K^+ conductance in cardiac Purkinje fibres

David C. Gadsby

Laboratory of Cardiac Physiology,
The Rockefeller University, 1230 York Ave., New York,
New York 10021, USA

Hormonal modulation of the ionic conductance of cell membranes is a topic of considerable current interest^{1–3}; it has a major role, for example, in the improved performance of the vertebrate heart elicited by sympathetic nerve stimulation or by circulating catecholamines, an effect involving enhanced calcium influx^{3–6}. β -Agonist catecholamines also abbreviate the action potential of cardiac Purkinje fibres^{7–11}, and increase the resting potential in a variety of cells^{12–14}, including cardiac cells^{7,10,15–17}, a hyperpolarization usually attributed to stimulation of the electrogenic Na^+/K^+ pump. We show here that nanomolar concentrations of β -catecholamines cause hyperpolarization of cardiac Purkinje fibres, not by increasing Na^+/K^+ pump current, but by increasing resting membrane K^+ conductance. The hyperpolarization and shortening of the action potential should increase availability of Na^+ channels and reduce the refractory period, effects tending to safeguard impulse propagation through the ventricular conducting system despite the increased heart rate caused by β -catecholamine action on the sinus node pacemaker.

Small bundles of Purkinje cells, 1–2 mm long and ≤ 0.2 mm wide (including connective tissue sheath), dissected from right ventricles of canine hearts, were suspended in a modified Hodgkin-Horowitz¹⁸ fast-flow chamber and voltage clamped using two microelectrodes. Currents were low-pass filtered (time constant, 100 ms). At a normal extracellular K^+ concentration ($[K^+]_o$) of 4 mM, in either normal chloride Tyrode's solution or low-chloride solution (major anion isethionate), the N-shaped, steady-state current-voltage relationship of these preparations often includes three zero-current potentials so that there are two possible levels of resting potential: the usual one near -90 mV and a 'lower' one near -40 mV¹⁹.

Figure 1A, a shows that, with the voltage-clamp amplifier disconnected, the resting potential can be switched from the lower to the higher level by injecting a small hyperpolarizing current pulse. After switching the resting potential back to the lower level with a depolarizing current pulse (not shown) the fibre was exposed to the β -adrenoceptor agonist, isoprenaline (100 nM; Fig. 1A, b), which caused, after a brief delay, a marked hyperpolarization with a complex time course. On washing out the isoprenaline the membrane potential declined slowly by about 2 mV and then remained at the higher resting level where subsequent application of isoprenaline resulted in a small reversible hyperpolarization (~ 2 mV, Fig. 1A, b). Although the initial

part of the hyperpolarization in Fig. 1A, b (near -40 mV) might be attributable to increased steady-state delayed rectifier current, this cannot be true of the later increases in resting potential, because the delayed rectifier is known to be inactivated negative to -50 mV^{8,11}.

The changes in membrane current underlying the hyperpolarizing response to β -adrenoceptor activation, recorded in a different fibre under voltage clamp, are shown in Fig. 1B. Similar results to those obtained with isoprenaline (Fig. 1B, a) could be elicited by noradrenaline when care was taken first to block α -adrenoceptors with $0.5 \mu M$ phentolamine (Fig. 1B b): thus, at holding potentials near the lower resting potential, after a delay of some 15 s the holding current shifted in the outward direction, reaching an early peak after about 30 s, then hesitating or declining slightly before rising more slowly to a later, broader peak (after about 2 min) from which a very slow decay occurred towards an eventual steady-state level, in the maintained presence of the catecholamine. The latter, slow decay reflects an apparent 'desensitization' or 'fatigue'²⁰ from which recovery (tested by repeated brief applications of catecholamine) is also slow, requiring many minutes of catecholamine-free superfusion. The records in Fig. 1B show that on washing out the catecholamines the current rapidly and monotonically returns to the control level. In some experiments, particularly at potentials positive to -40 mV, a small slowly-decaying undershoot of the control level of holding current was revealed after washing out the catecholamine (see, for example, Fig. 2A).

The Na^+/K^+ pump in cardiac cells is known to be electrogenic²¹ and even in resting Purkinje fibres it generates a measurable, steady, outward component of membrane current²². To test the hypothesis that the β -catecholamine-induced

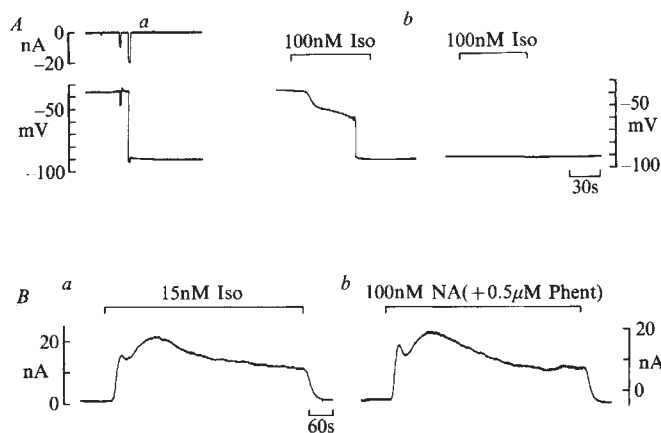


Fig. 1 β -Catecholamine-induced hyperpolarization (A) or outward current under voltage clamp (B). A, Changes in membrane potential caused by brief, 10- or 20-nA, hyperpolarizing current pulses (a), or by applications of 100 nM isoprenaline bitartrate (b), as indicated above the records. The two gaps in the voltage trace mark omission of 20- and 12-min sections of the record, respectively. The time calibration applies to both a and b. The low-chloride solution used in this experiment contained (in mM): K-methylsulphate, 4; Na isethionate, 137; $NaHCO_3$, 12; Camethanesulphonate, 2.7; NaH_2PO_4 , 1.8; $MgCl_2$, 0.5; dextrose, 5.5, and was equilibrated with 95% O_2 /5% CO_2 . All solutions in all experiments contained, in addition, 20 μM Na_2EDTA and 0.1 μM reduced glutathione to retard oxidation of the catecholamines²⁶. B, Changes in membrane current recorded at a steady holding potential of -45 mV in response to 8-min exposures to 15 nM isoprenaline HCl (a), or to 100 nM noradrenaline bitartrate (NA) in the presence of $0.5 \mu M$ phentolamine-HCl (Phent, b). Exposure to phentolamine, to block α -adrenoceptors, was begun 4 min before applying the noradrenaline. The chloride Tyrode's solution used in this experiment contained (in mM): KCl, 4; NaCl, 144; HEPES, 10 (adjusted to pH 7.3); $CaCl_2$, 2.7; NaH_2PO_4 , 1.8; $MgCl_2$, 0.5; dextrose, 5.5, and was equilibrated with oxygen.

outward current reflects increased Na^+/K^+ pump current, we recorded the response to isoprenaline at different membrane potentials, before and during inhibition of the pump by exposure either to a high concentration of the cardiac steroid acetylstrophanthidin or to K^+ -free fluid. The record in Fig. 2A shows first the large outward current caused by a control 40-s exposure to 100 nM isoprenaline. Application of $2\text{ }\mu\text{M}$ acetylstrophanthidin then caused a slow inward shift of holding

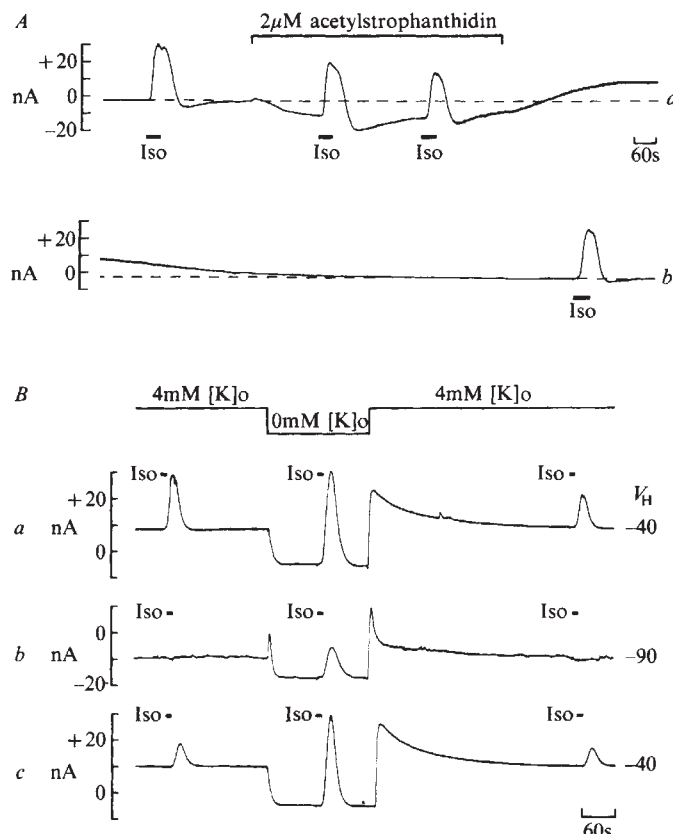


Fig. 2 Isoprenaline causes an outward shift of holding current even after inhibition of the Na^+/K^+ pump. **A**, Current changes caused by 40-s exposures to 100 nM isoprenaline (bars labelled Iso) at a holding potential of -30 mV ; **a** and **b** form a continuous record. The dashed line marks the control level of holding current. $2\text{ }\mu\text{M}$ acetylstrophanthidin was applied for the 12-min period indicated in **a**: test isoprenaline exposures were made after 3 min, and after 8 min, of acetylstrophanthidin action. Pump inhibition is demonstrated by the inward shift of holding current on application of acetylstrophanthidin as well as by the transient outward current on washing it off. The outward current caused by isoprenaline at the beginning of **a** was $\sim 20\%$ smaller than that elicited 7 min earlier by a similar isoprenaline exposure (not shown), presumably because of persisting 'desensitization'. Low-chloride solutions as in Fig. 1A. **B**, Current responses to 10-s exposures to 50 nM isoprenaline (bars labelled Iso) in 4 mM K^+ , and in K^+ -free, solution (indicated by the top line) first at a holding potential (V_H) of -40 mV (**a**), then -90 mV (**b**), and then again -40 mV (**c**). The interval between records **a** and **b** was 11 min, and that between **b** and **c** 5 min. The slow, exponentially declining overshoot of current recorded on switching from zero to 4 mM K^+ at -40 mV is a transient increment in pump current²². The more rapid transients recorded at -90 mV on switching to and from K^+ -free fluid are due to changes in K^+ current and reflect 'crossing over' of the steady-state, inwardly rectifying, K^+ current-voltage relationships at different $[\text{K}^+]_o$ levels so that, in these conditions, K^+ current is first increased and then decreased as K^+ concentration falls²³; an additional, slowly decaying increment in pump current is seen on switching back to 4 mM K^+ at -90 mV , but it is smaller than that recorded at -40 mV presumably because the Na^+ load (determined by the rate of Na^+ influx) is smaller at -90 mV than at -40 mV . The chloride Tyrode's solution used here contained $\text{HCO}_3^-/\text{CO}_2$ buffer.

current, reflecting the time course of pump inhibition and simultaneous abolition of steady outward pump current²². Despite that evident pump inhibition, test exposures to isoprenaline in the presence of the acetylstrophanthidin still induced a large outward current: the reduced size of the second isoprenaline response in acetylstrophanthidin probably reflects incomplete recovery from the desensitizing effect of the preceding response. The slowly rising and decaying transient overshoot of the control current level recorded on washing out the acetylstrophanthidin was presumably caused by a temporary increase in pump current, because dissociation of the cardiac steroid from pump sites must be expected to reveal temporary pump stimulation due to the raised intracellular Na^+ concentration ($[\text{Na}^+]_i$) which, in turn, resulted from the preceding pump inhibition²². The subsequent control exposure to isoprenaline, after washing out the acetylstrophanthidin for 16 min, again caused the usual outward current. The results in Fig. 2B demonstrate that the response to brief (10 s) applications of 50 nM isoprenaline remains undiminished during K^+ -free superfusion both at a holding potential of -40 mV and at -90 mV . The upper and lower current records (**a**, **c**) were obtained at -40 mV , and both show a large, exponentially decaying increment in pump current elicited by returning to 4 mM K^+ fluid after $\sim 3\text{ min}$ of exposure to K^+ -free solution. This pump current increment reflects temporary stimulation of the pump by the raised $[\text{Na}^+]_i$ resulting from pump inhibition during the preceding K^+ -free superfusion. In spite of that pump inhibition, it is clear that the size of the isoprenaline-induced outward current is not reduced in K^+ -free fluid. In fact, at both holding potentials the isoprenaline current is larger in the nominally K^+ -free solution than when $[\text{K}^+]_o$ is 4 mM . The outward current induced by isoprenaline both at -40 mV and at -90 mV is thus neither abolished nor markedly diminished in conditions in which the Na^+/K^+ pump is evidently inhibited, strongly suggesting that this catecholamine-induced outward current does not reflect enhanced Na^+/K^+ pump current.

Indeed, the voltage dependence and $[\text{K}^+]_o$ dependence of the isoprenaline current illustrated in Fig. 2B suggest that the underlying mechanism might be an increase in resting membrane K^+ conductance. Those results can be summarized by noting that: (1) at a given $[\text{K}^+]_o$, the isoprenaline-induced current becomes smaller when the holding potential (V_H) is shifted from -40 mV (upper and lower traces) to -90 mV (middle trace), that is, moved towards the presumed level of the K^+ equilibrium potential (E_K ; about -100 mV for 4 mM K^+) and considerably more negative for nominally zero $[\text{K}^+]_o$. (2) At both -40 and -90 mV , the isoprenaline-induced current becomes smaller when $[\text{K}^+]_o$ is raised from zero to 4 mM , that is, when E_K is moved towards the holding potential. In other words, the amplitude of the isoprenaline-induced current seems to vary with the expected size of the driving force for K^+ ions across the cell membrane (given by $V_H - E_K$). This relationship is demonstrated more clearly by the results in Fig. 3.

The traces in Fig. 3A show current changes induced by 20-s applications of 100 nM isoprenaline in a preparation held at various steady potential levels during superfusion with 16 mM K^+ solution. It is clear that, at 16 mM K^+ , the isoprenaline-induced current is large and outward at -40 mV , declines progressively as the holding potential is made more negative, becoming immeasurably small near -60 mV , and is plainly inward in direction at -65 mV . The current-voltage plot in Fig. 3B shows the steady levels of holding current (open circles) and the peak levels attained during the isoprenaline responses (filled circles), the vertical lines indicating the isoprenaline-induced current change. In Fig. 3C, the peak current changes (ΔI) are shown plotted against holding potential and it is evident that the isoprenaline-induced current varies roughly linearly with potential over the same 25-mV voltage range within which the steady-state, net current-voltage relationship (presumably dominated by 'background' K^+ current²³) shows the marked curvature characteristic of inward rectification. The least-squares regression line fitted to the points

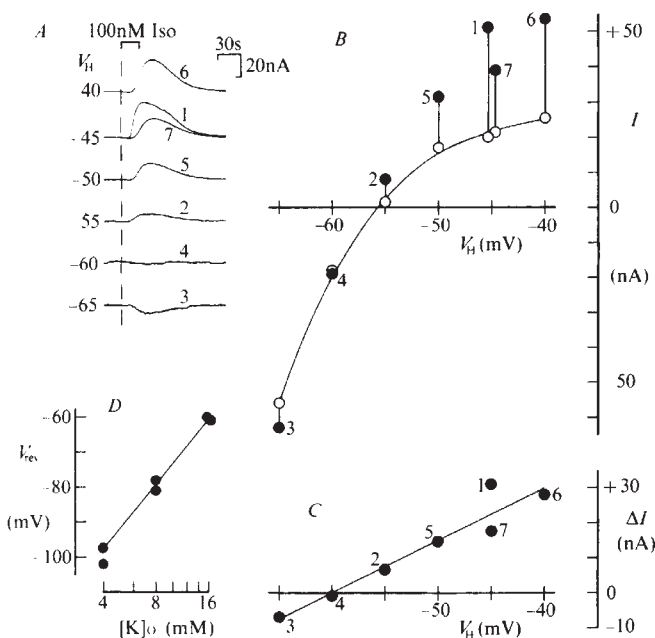


Fig. 3 Voltage dependence of isoprenaline-induced current. **A**, Current responses to 20-s exposures to 100 nM isoprenaline (Iso) in 16 mM K^+ , low-chloride solution (HCO_3^-/CO_2 buffer) at steady holding potentials (V_H) of -40 mV to -65 mV as indicated. Numbers 1-7 denote the recording sequence (also in **B** and **C**). Isoprenaline exposures were spaced ~6 min apart; the difference between the two responses at -45 mV presumably reflects cumulative 'desensitization' over the intervening ~40 min due to incomplete recovery between exposures. **B**, Graph of results in **A**. Abscissa, V_H ; ordinate, membrane current (I); open circles show steady holding currents, and filled circles show peak currents during responses to isoprenaline. The curve through the steady holding current was drawn by eye. **C**, Plot of peak, isoprenaline-induced current changes (ΔI), from **A** and **B**, against V_H . The straight line is a least-squares fit to all seven points and gives a reversal potential (V_{rev}) of -60 mV for ΔI . **D**, Semilogarithmic plot of V_{rev} against $[K^+]_0$; each point was obtained from a different fibre by linear interpolation of ΔI values recorded within 20 mV of V_{rev} (except for the most negative value at 4 mM $[K^+]_0$ which required a 2-mV linear extrapolation). The straight line shows E_K estimated for $[K^+]_i = 155$ mM. **A** and **D** are taken from ref. 27. Note that the waveform of the isoprenaline current remained relatively constant as V_H was varied (**A**); even the more complex time course of the current changes in Fig. 1B was well maintained for -90 mV $\leq V_H \leq$ -30 mV, implying that the shape reflects predominantly the time course of changes in K^+ conductance rather than the superposition of changes in inward and outward components of membrane current.

in Fig. 3C yields a reversal potential of -60 mV for the isoprenaline-induced current at 16 mM $[K^+]_0$. This reversal potential is included with five others, all determined in a similar way, in the semilogarithmic plot of Fig. 3D. The straight line, which provides a reasonable fit to the points, shows E_K calculated from the Nernst equation, $E_K = RT/F \ln ([K^+]_0/[K^+]_i)$, where R and F have their usual values, assuming that intracellular K^+ concentration ($[K^+]_i$) = 155 mM¹⁹, and $T = 37^\circ C$. These results confirm that the isoprenaline-induced current varies in size with the driving force on K ions, and reverses at E_K , and so argue forcefully that isoprenaline acts to increase steady-state membrane K^+ conductance. A similar β -adrenoceptor-mediated increase in resting K^+ conductance has recently been reported for atrial muscle cells of the canine coronary sinus²⁴ and, just as in the present experiments (For example, Fig. 2B, b, 0 mM $[K^+]_0$), the change in K^+ conductance could be elicited over a wide range of membrane potentials from plateau levels up to -100 mV. This finding strongly suggests that the present

catecholamine-modulated K^+ conductance is distinct from the delayed rectifier of cardiac Purkinje fibres, which Tsien *et al.*⁸ have previously shown to be augmented by β -catecholamines via a cyclic AMP-mediated pathway, because no steady-state flow of delayed rectifier current is expected at membrane potentials more negative than -40 to -50 mV^{8,11}. At voltages positive to -40 mV, the relative contributions of these two systems to the net catecholamine-induced increase in K^+ conductance is unclear.

Two of the present results indicate that the β -catecholamine-induced K^+ conductance lacks the marked inward rectification characteristic of the 'background' K^+ conductance of cardiac and skeletal muscle cells: (1) the current-voltage relationship is approximately linear over a limited voltage range (Fig. 3C), and (2) the isoprenaline-induced current is larger in zero $[K^+]_0$ than in 4 mM $[K^+]_0$, at both -90 mV and -40 mV (Fig. 2B), demonstrating that the corresponding current-voltage relationships do not cross each other, unlike those of the background K^+ conductance (see Fig. 2B legend). This lack of inward rectification implies that in physiological conditions the greatest effect of the catecholamine-induced K^+ conductance will be seen on the action potential plateau because the increment in outward current is larger (and membrane slope conductance is smaller) there than at diastolic potentials. This current must be expected to contribute to the abbreviation of the action potential seen in Purkinje fibres during β -catecholamine action.

The isoprenaline-induced current persists in the presence of acetylthiocholine, is augmented in K^+ -free fluid, and reverses close to presumed levels of E_K : these findings argue convincingly that the catecholamine-induced hyperpolarization studied here results from increased K^+ current and not from increased pump current. It seems possible that, in appropriate conditions, increased outward K^+ current might lead to extracellular K^+ accumulation and, hence, to an indirect stimulation of the Na^+/K^+ pump. To what extent such indirect effects might contribute to the enhanced Na^+/K^+ pump activity observed during catecholamine action^{15-17,25} remains an important question demanding further investigation.

I thank Dr Paul F. Cranefield for advice and encouragement throughout this work which was supported by USPHS grant HL-14899 and by an Established Fellowship of the New York Heart Association.

Received 9 May; accepted 17 October 1983.

1. Siegelbaum, S. A., Camardo, J. S. & Kandel, E. R. *Nature* **299**, 413-417 (1982).
2. Madison, D. V. & Nicoll, R. A. *Nature* **299**, 636-638 (1982).
3. Reuter, H., Stevens, C. F., Tsien, R. W. & Yellen, G. *Nature* **297**, 501-504 (1982).
4. Niederganger, R. & Page, S. *Proc. R. Soc. B* **197**, 333-362 (1977).
5. Reuter, H. *Rev. Physiol.* **41**, 413-424 (1979).
6. Osterrieder, W. *et al. Nature* **298**, 576-578 (1982).
7. Dudel, J. & Trautwein, W. *Experientia* **12**, 396-398 (1956).
8. Tsien, R. W., Giles, W. & Greengard, P. *Nature* **240**, 181-183 (1972).
9. Giotti, A., Ledda, F. & Mannaioni, P. F. *J. Physiol., Lond.* **229**, 99-113 (1973).
10. Trautwein, W. & Schmidt, R. F. *Pflügers Arch. ges. Physiol.* **271**, 715-726 (1960).
11. Kass, R. S. & Wiegand, S. E. *J. Physiol., Lond.* **322**, 541-558 (1982).
12. Clausen, T. & Flatman, J. A. *J. Physiol., Lond.* **270**, 383-414 (1977).
13. Somlyo, A. P. & Somlyo, A. V. *Fedn Proc.* **28**, 1634-1642 (1969).
14. Bulbring, E., Ohashi, H. & Tomita, T. in *Smooth Muscle: An Assessment of Current Knowledge* (eds Bulbring, E., Jones, A. W. & Tomita, T.) 219-248 (University of Texas Press, Austin, 1981).
15. Akasu, T., Ohta, Y. & Koketsu, K. *J. Heart J.* **18**, 860-866 (1977).
16. Vassalle, M. & Barnabei, P. *Pflügers Arch. ges. Physiol.* **322**, 287-303 (1971).
17. Wasserstrom, J. A., Schwartz, D. J. & Fozzard, H. A. *Am. J. Physiol.* **243**, H670-H675 (1982).
18. Hodgkin, A. L. & Horowitz, P. J. *Physiol., Lond.* **148**, 127-160 (1959).
19. Gadsby, D. C. & Cranefield, P. F. *J. gen. Physiol.* **70**, 725-746 (1977).
20. Marsh, J. D., Barry, W. H., Neer, E. J., Alexander, R. W. & Smith, T. W. *Circulation Res.* **47**, 493-501 (1980).
21. Glitsch, H. G. *Rev. Physiol.* **44**, 389-400 (1982).
22. Gadsby, D. C. & Cranefield, P. F. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1783-1787 (1979).
23. Noble, D. *J. cell. comp. Physiol.* **66**, 127-135 (1965).
24. Boyden, P. A., Cranefield, P. F. & Gadsby, D. C. *J. Physiol., Lond.* **339**, 185-206 (1983).
25. Lee, C. O. & Vassalle, M. *Am. J. Physiol.* **244**, C110-C114 (1983).
26. Furchgott, R. F. *Pharmac. Rev.* **7**, 183-265 (1955).
27. Cranefield, P. F. & Gadsby, D. C. *J. Physiol., Lond.* **318**, 34P-35P (1981).

Chemical basis of antigenic variation in foot-and-mouth disease virus

D. J. Rowlands*, B. E. Clarke*, A. R. Carroll*,
F. Brown*, B. H. Nicholson†, J. L. Bittle‡,
R. A. Houghten‡ & R. A. Lerner‡

* Biochemistry Department, Animal Virus Research Institute, Pirbright, Woking, Surrey GU24 0NF, UK

† Department of Physiology and Biochemistry, The University, Whiteknights, Reading, Berkshire RG2 2AJ, UK

‡ Research Institute of Scripps Clinic, North Torrey Pines Road, La Jolla, California 92037, USA

One of the difficulties in controlling foot and mouth disease by vaccination is the occurrence of the virus as seven distinct serotypes because immunity conferred by vaccination against one serotype leaves the animals susceptible to infection by the other six. Moreover, the antigenic variation, even within a serotype, can be so great that immunity against the homologous strain of virus need not necessarily ensure protection against infection by other viruses within that serotype. Here we report the separation of three natural antigenic variants, distinguishable in cross-neutralization tests from an isolate of foot-and-mouth disease virus (FMDV). The serological differences could also be demonstrated by antisera elicited by synthetic peptides corresponding to residues 141–160 of the capsid polypeptide VP1, showing that this region contains a major immunogenic site of the virus. The results have practical implications for the choice of viruses for vaccine production.

There is considerable evidence that the major immunogen is situated on VP1, one of the four capsid proteins^{1–5}. Furthermore, the position of the major immunogenic site on VP1 of virus of serotype 0 has been shown to lie within the amino acid tract 141–160^{6–8}. The same region of VP1 from other serotypes also possesses immunogenic activity (J.L.B. *et al.*, unpublished data). Accumulation of sequence data for this region has allowed us to attempt to correlate the variability of the amino acid sequence between serotypes and subtypes with their serological relationship (manuscript in preparation).

Antisera elicited by peptide 141–160 of virus of serotype A, subtype 12, strain 119, predicted from the nucleotide sequence determined by Kleid *et al.*⁹ neutralized our stocks of this virus only poorly. This contrasted with our results with sera produced against synthetic peptides of viruses of serotypes 0, subtype 1⁷, A, subtype 10¹⁰ and C, subtype 3¹¹. However, the virus from which the nucleotide sequence had been determined by molecular cloning methods⁹ (provided by Dr D. M. Moore, Plum Island Animal Disease Center and referred to here as A12 USA) was neutralized well. The antigenic characteristics of FMDV strains produced in cell cultures may be influenced by the passage history of the virus^{12,13}. Our A12 virus (A12 AVRI), the parent of that used by Kleid *et al.*⁹, had been passaged only once in BHK cells without plaquing, in contrast to the multiple passaging and plaquing referred to by Kleid *et al.*

A synthetic oligonucleotide, complementary to a conserved sequence^{10,14} downstream from VP1 and having the sequence dACGTCACCCGCCCACTT (provided by Drs H. Mathes and M. J. Gait, MRC, Molecular Biology Laboratory, Cambridge) was used for primer extension sequencing of the USA and AVRI viruses^{15,16}. The oligonucleotide was labelled with ³²P and used to direct the synthesis of cDNA transcripts by reverse transcriptase. The nucleotide sequence of these primer extended fragments was then determined by standard methods^{17,18}.

The nucleotide sequence obtained for the A12 USA virus agreed exactly with that determined by molecular cloning methods⁹, except that we found an extra Leu codon at the C

Table 1 Cross neutralization of viruses containing different amino acids at positions 148 and 153 of VP1 with homologous and heterologous antiviral and anti-peptide sera

		log ₁₀ virus neutralized by 0.015 ml serum dilution					
		Serum dilution	148 153	A Ser Leu	B Leu Pro	C Ser Ser	USA Phe Pro
A							
Ser at 148		1/2		5.5	3.7	4.1	3.9
Leu at 153	Anti-virion	1/20		4.3	3.3	3.5	2.9
		1/200		3.3	1.3	1.7	0.7
		1/2,000		1.5	0.1	1.3	0.3
	Anti-peptide	1/1		2.5	1.1	1.5	0.9
B							
Leu at 148		1/2		4.5	5.3	3.7	4.3
Pro at 153	Anti-virion	1/20		3.7	4.7	2.9	3.5
		1/200		0.9	3.7	0.9	2.1
		1/2,000		0.7	1.7	0.9	0.1
	Anti-peptide	1/1		Nil	3.2	0.5	2.2
C							
Ser at 148		1/2		5.7	5.3	7.3	4.9
Ser at 153	Anti-virion	1/20		4.9	3.7	4.5	3.3
		1/200		3.5	2.9	4.1	2.5
		1/2,000		1.9	0.5	2.5	0.5
	Anti-peptide	1/1		1.5	0.1	2.6	1.5
USA							
Phe at 148		1/2		4.5	3.8	4.1	5.6
Pro at 153	Anti-virion	1/20		3.9	3.5	4.1	4.7
		1/200		1.3	2.5	1.1	3.7
		1/2,000		1.3	0.5	0.5	2.5
	Anti-peptide	1/1		Nil	2.0	Nil	2.7
Leu at 148	Anti-peptide	1/1		0.3	0.9	0.5	0.7
Leu at 153							
Phe at 148	Anti-peptide	1/1		Nil	0.4	0.4	0.3
Leu at 153							

Tenfold dilutions of each virus were mixed with an equal volume of normal serum or antiserum dilution and 0.03 ml of each mixture injected intraperitoneally into groups of 7 days old mice. The values given are the difference between the titre of the virus plus normal serum and the virus plus antiserum dilution.

terminus which was not recorded by Kleid *et al.*⁹ but which was found on re-examination of the sequencing gels (D. M. Moore, personal communication). However, the nucleotide sequence of the A12 AVRI virus was ambiguous at the positions representing amino acids 148 and 153 of VP1, suggesting the presence of more than one strain of virus in the original stock (Fig. 1). Our stock virus was therefore plaqued on BHK 21 cell monolayers, ten separate clones grown, their RNA extracted and primer extension sequencing reactions performed. Of the ten viruses, three had Ser at 148 and Leu at 153 (virus A), five had Leu at 148 and Pro at 153 (virus B), and two had Ser at both 148 and 153 (virus C), whereas the A12 USA virus had Phe at 148 and Pro at 153. In addition, all ten clones had Val at 171 compared with Glu at this position in the USA virus.

Sera prepared by injection into guinea pigs of vaccines prepared from the USA and the three variant viruses were tested for their neutralizing activity against the homologous and heterologous viruses. The homologous viruses were neutralized to a much greater extent than the heterologous viruses (Table 1). The viruses fell into two distinct pairs, A and C, and B and USA, but even within these pairs, neutralization by the homologous serum was significantly greater than that by the heterologous serum.

Peptides corresponding to the 141–160 region of VP1 of each virus were synthesized chemically, linked to keyhole limpet haemocyanin with *N*-maleimidobenzoyl-*N*-hydroxysuccinimide ester (MBS) using an added Cys residue at the C terminus and injected subcutaneously into guinea pigs after mixing with Al(OH)₃ as adjuvant. Sera taken 21 days after injection neutralized the homologous viruses to a much greater extent than the heterologous viruses, emphasising the importance of amino acid substitutions at positions 148 and 153 (Table 1). Moreover, the viruses could again be grouped into A and C, and B and USA on the basis of these tests. The ratio between

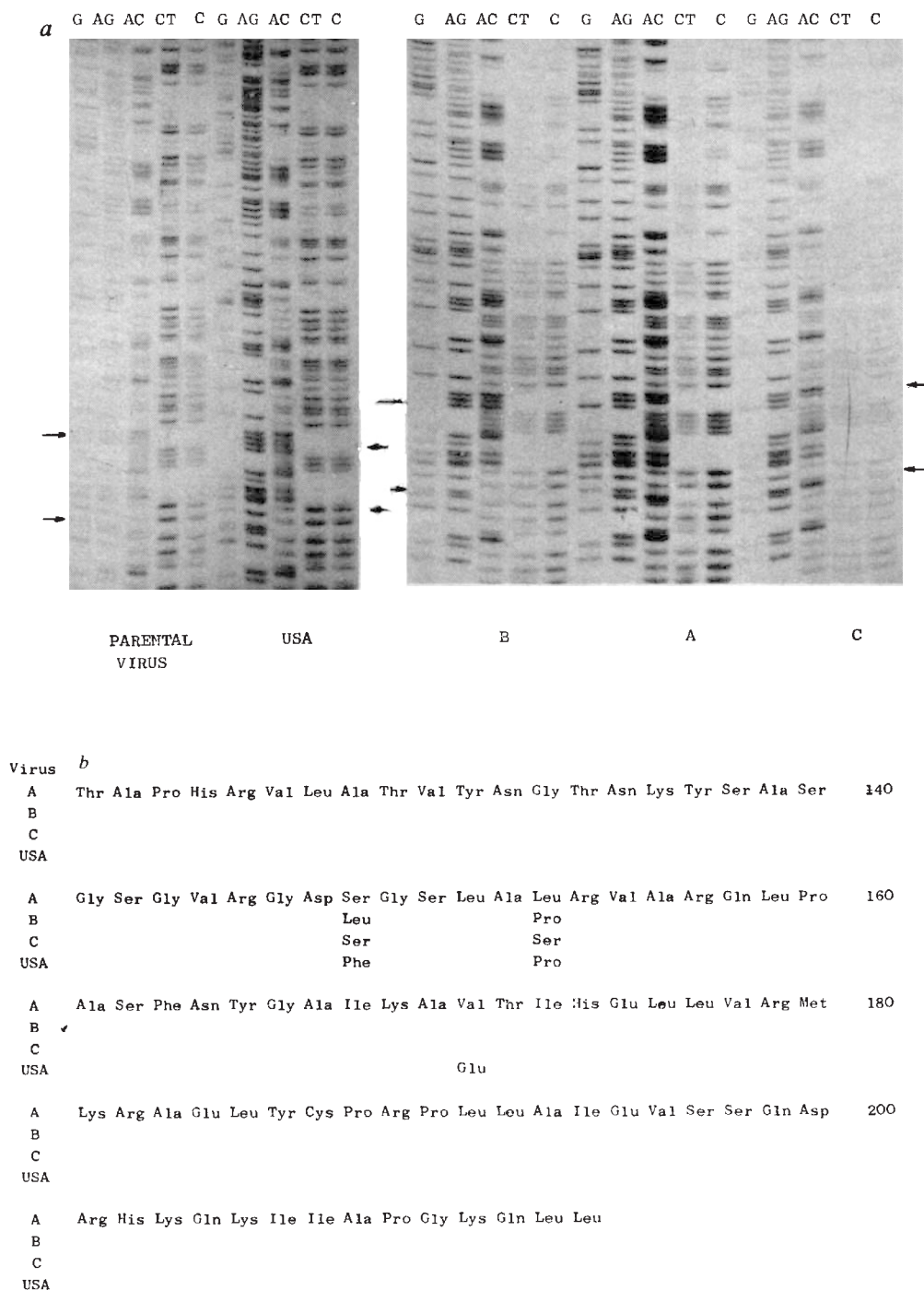


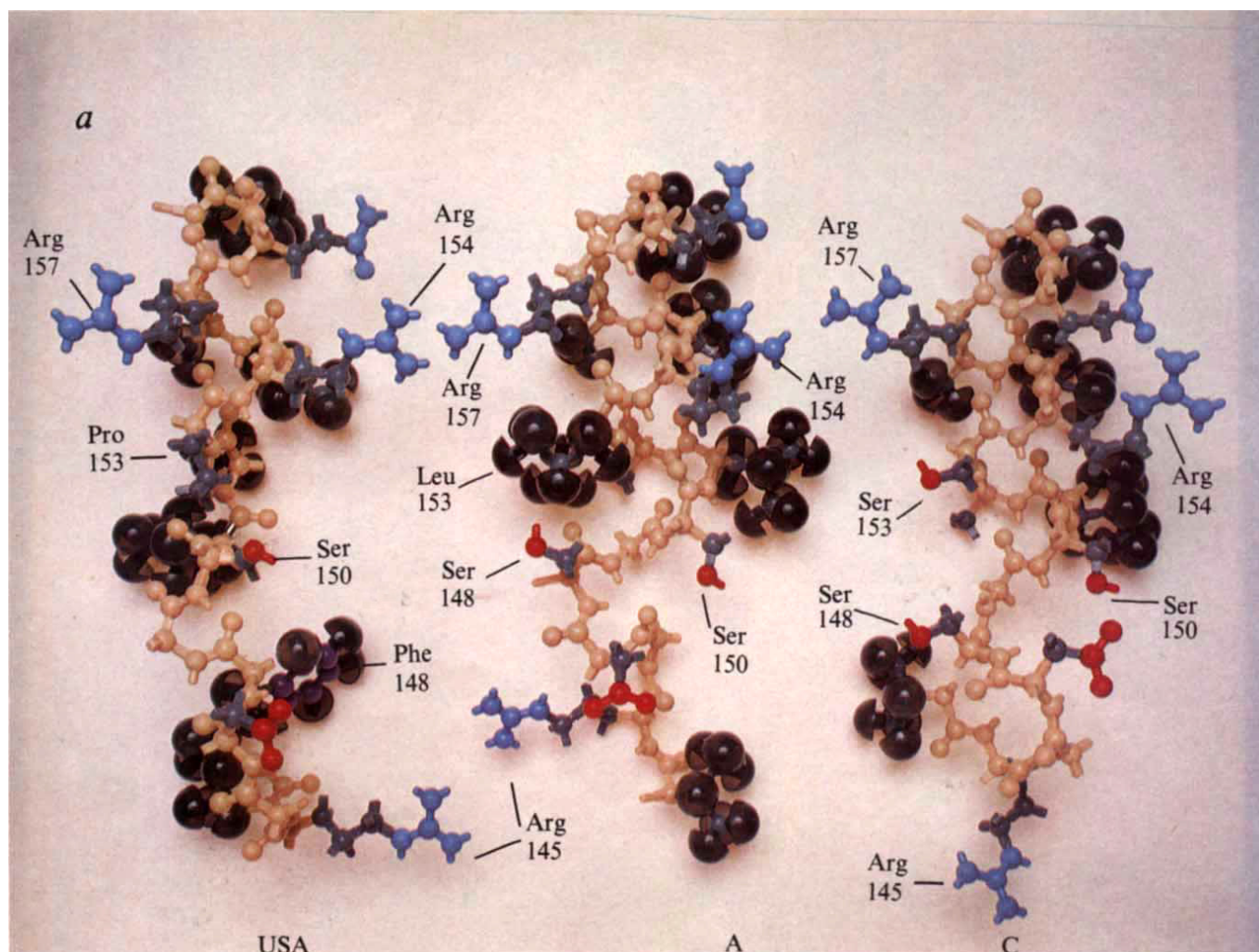
Fig. 1 Sequence gels of the DNAs corresponding to the C-terminal region of VP1 from variants of foot-and-mouth disease virus, serotype A, subtype 12. A synthetic DNA primer, complementary to a conserved sequence of the RNA at the 3' side of the VP1 coding region, was labelled with ^{32}P using polynucleotide kinase and 50 ng were hybridised to 20 μg of the appropriate virus RNA. The primer was extended using reverse transcriptase and size-selected on sucrose gradients after hydrolysing the RNA template in 0.1 M NaOH. The cDNA was sequenced by the method of Maxam and Gilbert^{17,18} and the fragments separated by electrophoresis on 6% polyacrylamide gels. Panel **a** shows a portion of such a gel corresponding to the nucleotide sequence coding for the major antigenic region. The positions at which the sequences of the variants differ are indicated by arrows. At the two positions marked on the parental stock of the virus it is impossible to determine whether the nucleotides are A or G. The predicted amino acid sequences for position 121 to the C terminal end of VP1 are given in panel **b**. The amino acid numbering system is normalized to that for serotype 0-1 (Kaufmann) given in ref. 14.

the neutralization indices of the antipeptide and the antiviral sera with the homologous and heterologous viruses is similar, suggesting that the amino acid sequence at 141–160 contains the major antigenic determinant of the virus particle.

Peptides with Leu at 148 and 153 and with Phe at 148 and Leu at 153 had been synthesized before it was realized that our stock virus consisted of a mixture of viruses. Antisera against these two peptides did not neutralize any of the four viruses, again emphasising the importance of residues 148 and 153. Thus, for position 148, virus A with Ser at 148 and Leu at 153,

is not neutralized by anti Leu-Leu or anti Phe-Leu sera. Similarly, for position 153 anti Leu-Leu serum did not neutralize virus B (Leu Pro) and anti Phe-Leu serum did not neutralize the USA virus (Phe Pro).

A virus with a sequence at 141–160 identical to that used by Kleid *et al.*⁹ was not isolated, presumably because it represented such a small proportion of the viruses in our parent stock, or because it had arisen by mutation during the many passages it has undergone in the USA laboratory. The Val at 171 of the three viruses in our parental stock compared to Glu in the USA



<i>b</i> Virus	144										150										155										159									
USA	Val+	Arg	Gly	Asp	Phe+	Gly	Ser	Leu+	Ala+	Pro	Arg	Val+	Ala+	Arg	Gln	Leu+																								
A	+				Ser		→	+	-	Leu+		+	+																											
B					Leu+			+	+	Pro		+	+																											
C	+				Ser		→	+	+	Ser		+	+																											

Fig. 2 Core residue predictions and molecular models of the 144-159 region of variants of FMDV. The sequences in *b* show the substitutions in viruses A, B and C compared to the USA virus. → indicates the N terminus of the helix which is terminated at its carboxyl end by Pro 160. Hydrophobic residues predicted in core are indicated by + and non-core hydrophilic residues by -. The helix in the USA virus runs from Leu 159 to Pro 153. In viruses A and C, the helix is nearly a turn longer, being terminated at residue 150 by Gly 149. Virus B is not shown since it differs from the USA virus only by the substitution of Leu for Phe at 148. The side chains of the residues predicted in core are shown with full van der Waals' radii for ease of identification. All appear on the lower side of the helix and are unavailable for antibody recognition. The course of the chain near the N terminus is uncertain but is restricted by the need for the predicted core residues to be underneath and for the N-terminus to connect with the rest of the protein. Because of the difference in the length of helices in the USA virus and virus B on the one hand and viruses A and C on the other, they must adopt different conformations and thus the position of the side chains (for example, Asp 147, Ser 150) will be different in these two classes.

virus was considered to be unrelated to antigenic specificity since it was a constant feature of the three antigenically distinct variants. It is presumed to lie outside the antigenic region⁶⁻⁸.

The specificity of the virus and peptide antisera was also investigated by radioimmune precipitation. The virus antisera did not discriminate between homologous and heterologous viruses (Table 2) presumably due to the presence of common antigenic determinants on the viruses which are not of importance in the neutralization of infectivity (Fig. 2). Peptide antisera, however, were highly specific and clearly differentiated the viruses into two classes with related antigenic characteristics which corresponded to the groupings revealed by the neutralization tests (Tables 1 and 2).

The four viruses also fall into the same two groupings by the independent criteria of the predicted molecular structures of

their antigenic regions^{19,20}. Many of the side chains are unlikely to be available for antibody recognition, particularly those that are buried in the hydrophobic region of the protein and thereby prevented from acting as determinants. It is useful to identify such side chains because of the restraint they place on the main chain atoms. The results of the core prediction²⁰ for the antigenic region of the four viruses are shown in Fig. 2. The secondary structure prediction of Chou and Fasman¹⁹ shows that considerable lengths of the sequence are expected to be helical, although more than a turn shorter than for virus of serotype 0, subtype 1 for which a helix had been predicted previously⁸. The core and secondary structure predictions have been combined to build a theoretical model of the antigen determinant site (Fig. 2).

The USA virus and virus B differ only in one residue. The small variation that is observed in the neutralization tests is

Table 2 Immune precipitation of ^{35}S -labelled viruses containing different amino acids at positions 148 and 153 of VP1 with homologous and heterologous antiviral and antipeptide sera

Serum	Serum dilution	Percentage of ^{35}S virus precipitated by 20 μl of serum dilution			
		A Ser Leu	B Leu Pro	C Ser Ser	USA Phe Pro
Antivirus A	1/1	100	74	87	116
	1/10	32	32	27	41
Antipeptide A	1/1	92	0	100	0
	1/2	49	0	53	0
Antivirus B	1/1	100	100	83	176
	1/10	40	33	43	38
Antipeptide B	1/1	17	100	17	73
	1/2	11	42	13	36
Antivirus C	1/1	100	100	100	100
	1/10	31	25	30	52
Antipeptide C	1/1	95	5	100	3
	1/2	55	3	66	1
Antivirus USA	1/1	42	51	39	100
	1/10	18	17	18	56
Antipeptide USA	1/1	20	97	4	98
	1/2	8	84	2	100

Radiolabelled preparations of viruses A, B, C and USA were made by growing them in the presence of ^{35}S -methionine and purifying by differential centrifugation and sucrose density gradient separation. The viruses were diluted in TNEN (50 mM Tris-HCl, pH 7.5, 0.2 M NaCl, 5 mM EDTA, 0.05% Nonidet NP40) to give approximately the same ^{35}S content (c. 20,000 d.p.m. ml^{-1}). One ml aliquots of virus were mixed with 20 μl of serum dilution (in TNEN) and incubated for 24 h at 4 °C. *Staphylococcus aureus* ghosts which had been washed three times in TNEN were added to the virus serum mixtures using a constant ratio of *S. aureus* ghosts to serum. After 30 mins at 20 °C the *S. aureus* ghosts were centrifuged, washed with TNEN and the final pellets dissolved in tissue solubiliser solution (NCS) and counted by liquid scintillation. The results are expressed as percentages of the maximum number of counts adsorbed to the *S. aureus* ghosts by the homologous antiserum.

probably due to the Phe-Leu residues at position 148, which are predicted to be in core. These could influence sequentially adjacent residues by adopting different dihedral angles within their similar permitted range²¹, thereby causing some change in the spatial positions of the neighbouring exposed groups. These two viruses have a very short helix, terminating at Pro 153. In contrast, viruses A and C have longer helices (Fig. 2) and this is reflected by the neutralization data (Table 1).

A further distinction between the two groups of viruses is the substitution of the core-predicted Leu or Phe at 148 by Ser, which removes the restraint for the main chain to be near the core. This means that the non-helical sequence 145–150 of viruses A and C is free to 'loop out', so that the hydrophilic groups in this region could well occupy distinctive spatial positions. The difference between viruses A and C lies in the substitution of Ser in virus C for Leu in virus A at 153. The polar hydroxyl of the Ser is near the top side of the helix (Fig. 2) and its substitution by a non-polar Leu might be expected to cause a large difference in neutralization tests. In fact it does not, suggesting that the Leu side chain points to the underside of the helix, and this is confirmed by the core prediction for this residue. Thus the difference between viruses A and C effectively reduces to the presence or absence of the polar hydroxyl group and they are no more dissimilar from each other than are virus B and the USA virus (Tables 1 and 2).

Our demonstration that a virus sample from the natural bovine host, passaged only once in tissue culture cells, is composed of approximately equal numbers of three variants which differ in sequence in the antigenic site region contrasts strongly with the reports of the selection of mutants under laboratory conditions by host range²² or immunological pressure²³. In the latter case the frequency of the selection of variants is of the order of 10^{-4} .

The antigenic variability described here with viruses that were thought to be the same indicates that the amino acid sequences of the antigenic sites of viruses being used for vaccine production ideally should be the same as those of the field isolates. Con-

sequently a greater measure of control is required to ensure that the antigens used for vaccine production are the same as those of viruses occurring in the field.

Received 27 June; accepted 6 October 1983.

- Wild, T. F., Burroughs, J. N. & Brown, F. J. *gen. Virol.* **4**, 313–320 (1969).
- Laporte, J., Grosclaude, J., Wantyghem, J., Bernard, S. & Rouze, P. C. *r. hebdo. Séanc. Acad. Sci., Paris* **276**, 3399–3401 (1973).
- Bachrach, H. L., Moore, D. M., McKercher, P. D. & Polatnick, J. J. *Immun.* **115**, 1636–1641 (1975).
- Kaaden, O. R., Adam, K.-H. & Strohmaier, K. J. *gen. Virol.* **34**, 397–400 (1977).
- Meloen, R. H., Rowlands, D. J. & Brown, F. J. *gen. Virol.* **45**, 761–763 (1979).
- Strohmaier, K., Franze, R. & Adam, K.-H. J. *gen. Virol.* **59**, 295–306 (1982).
- Bittle, J. L. *et al. Nature* **298**, 30–33 (1982).
- Pfaff, E., Musgay, M., Bohm, H. O., Schulz, G. E. & Schaller, H. *EMBO J.* **1**, 869–874 (1982).
- Kleid, D. G. *et al. Science* **214**, 1125–1129 (1981).
- Boothroyd, J. C., Harris, T. J. R., Rowlands, D. J. & Lowe, P. A. *Gene* **17**, 153–161 (1982).
- Makoff, A. J., Paynter, C. A., Rowlands, D. J. & Boothroyd, J. C. *Nucleic Acids Res.* **10**, 8285–8295 (1982).
- Cowan, K. M., Erol, N. & Whiteland, A. P. *Bull. Off. Int. Epiz.* **81**, 1271–1298 (1974).
- Cowan, K. M., Erol, N. & Whiteland, A. P. J. *gen. Virol.* **41**, 437–442 (1978).
- Kurz, C., Forss, S., Kupper, K., Strohmaier, K. & Schaller, H. *Nucleic Acids Res.* **9**, 1919–1931 (1981).
- Ghosh, P. K., Reddy, V. B., Piatak, M., Lebowitz, P. & Weissman, S. *Meth. Enzym.* **65**, 580–585 (1980).
- Busslinger, M., Moschonas, N. & Flavell, R. A. *Cell* **27**, 289–298 (1981).
- Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, 499–560 (1980).
- Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 989–993 (1977).
- Chou, P. Y. & Fasman, G. D. *Adv. Enzym.* **47**, 45–148 (1978).
- Nicholson, B. H. *Biochem. Soc. Trans.* **10**, 387–388 (1982).
- Anfinsen, C. B. & Scheraga, H. A. *Adv. Protein Chem.* **29**, 205–300 (1975).
- Minor, P. D. *et al. Nature* **301**, 674–679 (1983).
- Schild, G. C., Oxford, J. S., De Jong, J. C. & Webster, R. G. *Nature* **303**, 706–609 (1983).

Mimicking the alloantigenicity of proteins with chemically synthesized peptides differing in single amino acids

Hannah Alexander*, David A. Johnson*, Jonathan Rosen†, Libuse Jerabek†, Nicola Green*, Irving Lerner Weissman† & Richard A. Lerner*

* Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037, USA

† Laboratory of Experimental Oncology, Department of Pathology, School of Medicine, Stanford University, Stanford, California 94305, USA

Recent studies have shown that short chemically synthesized peptides very often induce antibodies which react with the cognate sequence in the intact folded protein. Since such antibodies react with known regions of proteins, they are of pre-determined specificity and offer a precision not previously possible with immunological probes¹. A basic concept emerging from the use of such antibodies in viral systems is that the differential immunogenicity of closely related proteins can be mimicked by short peptides which span the regions of sequence variation^{2,3}. To generalize this concept, we have studied the two Thy-1 proteins which vary by only a single amino acid. Chemically synthesized peptides differing in only one out of 19 amino acids were able to induce allospecific antisera. Thus, single amino acid changes have similar effects on the immunogenicity of proteins and small peptides, even though the latter are free from constraints provided by neighbouring structures in the tertiary configuration of the intact folded proteins.

The Thy-1 glycoprotein exists in two allotypic forms, Thy 1.1 and Thy 1.2 (ref. 4). It has been proposed that an arginine residue at position 89 corresponds to the Thy 1.1 allotype and a glutamine residue to Thy 1.2 (ref. 5). Since the two proteins differ by only a single residue, they offered a good opportunity for testing the discriminatory power of antibodies induced by synthetic immunogens.

Based on the protein sequence⁵, we selected for synthesis a set of six peptides covering most of the Thy-1 molecule (Table

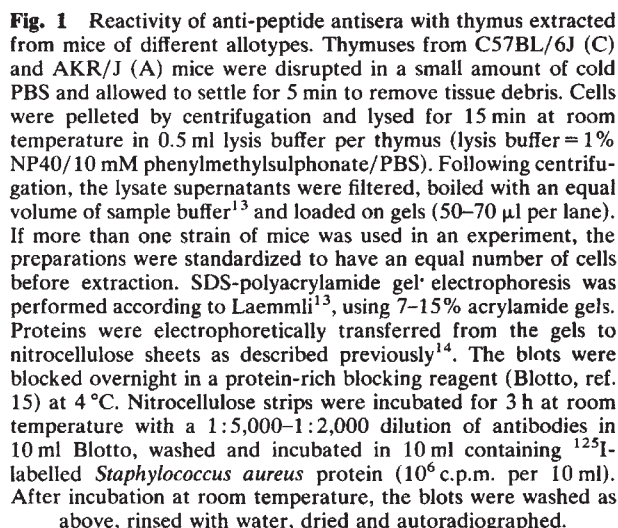


Fig. 1 Reactivity of anti-peptide antisera with thymus extracted from mice of different allotypes. Thymuses from C57BL/6J (C) and AKR/J (A) mice were disrupted in a small amount of cold PBS and allowed to settle for 5 min to remove tissue debris. Cells were pelleted by centrifugation and lysed for 15 min at room temperature in 0.5 ml lysis buffer per thymus (lysis buffer = 1% NP40/10 mM phenylmethylsulphonate/PBS). Following centrifugation, the lysate supernatants were filtered, boiled with an equal volume of sample buffer¹³ and loaded on gels (50–70 μ l per lane). If more than one strain of mice was used in an experiment, the preparations were standardized to have an equal number of cells before extraction. SDS-polyacrylamide gel¹ electrophoresis was performed according to Laemmli¹³, using 7–15% acrylamide gels. Proteins were electrophoretically transferred from the gels to nitrocellulose sheets as described previously¹⁴. The blots were blocked overnight in a protein-rich blocking reagent (Blotto, ref. 15) at 4°C. Nitrocellulose strips were incubated for 3 h at room temperature with a 1:5,000–1:2,000 dilution of antibodies in 10 ml Blotto, washed and incubated in 10 ml containing ¹²⁵I-labelled *Staphylococcus aureus* protein (10⁶ c.p.m. per 10 ml). After incubation at room temperature, the blots were washed as above, rinsed with water, dried and autoradiographed.

The anti-peptide antisera were tested for their ability to react with, and discriminate between Nonidet P-40 (NP40) extracts of thymus tissues from AKR/J (Thy 1.1) and C57BL/6J (Thy 1.2) mice. Antisera against peptides 85 and 87 failed to recognize the Thy-1 protein from either mouse strain. The inability of these antibodies to bind to the protein is not due to steric hindrance by glycan residues as recently proposed by other investigators⁵, since the removal of the sugar by endoglycosidase F⁶ did not render the molecule more accessible to the anti-peptide antibodies (data not shown). Antibodies to peptide 86 were strongly and equally reactive with both thymus extracts (Figs 1, 2). Antibodies to peptide 89 were of lower titre but also reacted with thymus extracts from both strains (data not shown). The results support the prediction that peptides 86 and 89 represent conserved regions of the mouse Thy-1 molecule. Antisera to peptides 88 and 90, which were predicted to be from allospecific regions of Thy 1.2 and 1.1 respectively, showed the corresponding preference for the C57BL/6J(C) Thy 1.2 and AKR/J(A) Thy 1.1 extracts (Fig. 1). Clearly, the presence of

Anti-peptide antiserum	Sequence position	Peptide antigen	Antibody titres
85	19-33	85	320-640
86	38-56	86	>1,280
87	57-72	87	640-1,280
88	79-98	88	320-640
		90	40-80
89	100-112	89	320-640
90	79-98	90	320-640
		88	160-320

10 20 30 40 50 60
 ZKVTSLTACLVNQNLRLDCRHNNTKDNDSIQHEFSLTREKRKHVLSGTLGPEHTYRSRV
 65 85 87
 70 80 90 100 110
 TLSNGPYIKVLTLANFTTKDEGDYFCQLGVSGANPMSSNYISIVYRDKLVKC
 87 88 89
 on R

The peptides used in the present studies are of such size that each could contain multiple immunogenic sites. Indeed, peptides

Table 2 Biological activity of anti-peptide antisera

Antisera	Membrane fluorescence*		Cytolysis†	
	Targets: AKR/J	C57BL/6J	AKR/J	C57BL/6J
	Thy 1.1	Thy 1.2	Thy 1.1	Thy 1.2
Anti-peptide 88	14	74	5	45
Anti-peptide 90	85	75	50	50
Rabbit anti-Thy-1	81	94	80	80
Preimmune rabbit	13	3	—	—

* One million suspended thymocytes were reacted for 1 h on ice with 75 μ l of a 1:10 dilution of anti-peptide antisera, preimmune serum or commercial rabbit anti-Thy-1 (Cappel). After three washes in cold RPMI 1640 plus 10% heat-inactivated fetal bovine serum, the cells were reacted with fluorescein-conjugated goat anti-rabbit IgA, G, M antibody (Cappel Labs), washed as above and fixed with 1% formalin in phosphate-buffered saline (PBS, pH 7.2). The % of fluorescent cells was determined with a FACS IV (Becton-Dickinson)¹¹.

† One million suspended thymocytes were reacted at 37°C with 100 μ l of a 1:10 dilution of heat-inactivated antiserum for 1 h. After three washes in RPMI 1640 plus 10% heat-inactivated fetal bovine serum, the cells were reacted for 1 h with a 1:20 dilution of Low-Tox rabbit complement (Accurate Chemical and Scientific) at 37°C. The % of dead cells was scored visually, following the addition of Trypan blue stain. Values are expressed ($\pm$ %) as the per cent cytolysis minus the % killed with preimmune rabbit serum with complement¹².

88 and 90 have 18 amino acids in common and differ with respect to only one. Nevertheless, the presence of a single distinguishing amino acid is sufficient to make a significant portion of the immune response specific for each peptide, resulting in preferential reactivity of the antisera with their inducing peptides as well as the intact protein.

Antibodies against peptide 88 have minimal cross-reactivity with the AKR/J Thy 1.1 molecule whereas peptide 90 seems to induce a broader, less specific, spectrum of antibodies. Thus, amino acid changes can either render or abolish differential immunogenicity. In the present study, replacement of Gln by Arg at position 89 diminishes differential immunogenicity and the resultant peptide appears now to be composed of a collection of more equally functioning antigenic determinants. A similar phenomenon has been observed in our studies of synthetic peptides which correspond to an antigenic determinant of the hepatitis B surface antigen⁸.

The important question is why should a single amino acid change similarly affect the immunogenicity of a peptide in solution and the cognate structure in a folded protein, where its

conformational flexibility is likely to be constrained by bonds with neighbouring structures. One possibility is that the peptide in solution has a favoured conformation which closely resembles that of the folded protein. This is, however, inconsistent with the body of evidence which suggests that, in aqueous environments, even relatively long peptides have very little ordered structure (reviewed in ref. 9). The answer to this question will require further studies, but several models involving induced fit and/or stabilization of regions of high mobility in proteins have already been discussed by us^{1-3,9,10}. Regardless of which model turns out to be correct, it already appears that studies using synthetic peptide immunogens will help unravel the rules governing the phenomenon of immunodominance as well as allow insight into the structure of proteins in solution.

We thank Susie M. Seaver and A. Lynlee Austell for technical assistance and Dr J. Elder for a gift of endoglycosidase F. This research was supported in part by USPHS grant 5 PO1 CA 25803 and AI 09072. D.A.J. is supported by USPHS grant CA 32630 from NCI. J.R. is supported by a postdoctoral fellowship from the Cancer Research Institute. This is publication no. 3106 MB from the Research Institute of Scripps Clinic.

Received 25 July; accepted 3 October 1983.

1. Lerner, R. A. *Nature* **299**, 592-596 (1982).
2. Bittle, J. L. *et al.* *Nature* **298**, 30-33 (1982).
3. Gerin, J. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 2365-2369 (1983).
4. Reif, A. F. & Allen, J. M. V. *J. exp. Med.* **120**, 413-433 (1964).
5. Williams, A. F. & Gagnon, J. *Science* **216**, 696-703 (1982).
6. Elder, J. H. & Alexander, S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4540-4544 (1982).
7. Barclay, A. N., Letarte-Muirhead, M. & Williams, A. F. *Nature* **263**, 563-567 (1976).
8. Alexander, H. & Lerner, R. A. *Advances in Hepatitis Research* (ed. Chisari, F. E.) (Masson, New York, in the press).
9. Niman, H. L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **80**, 4949-4953 (1983).
10. Green, N. *et al.* *Cell* **28**, 477-487 (1982).
11. Coons, A. H., Leduc, E. & Connolly, J. J. *exp. Med.* **102**, 49-55 (1955).
12. Boyse, E. A., Old, L. & Stockert, E. *Ann. N. Y. Acad. Sci.* **99**, 574-579 (1962).
13. Laemmli, U. K. *Nature* **227**, 680 (1970).
14. Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350-4354 (1979).
15. Johnson, D. A., Gautsch, J. W., Sportsman, J. R. & Elder, J. H. *Gene analyt. Tech.* (in the press).

Instability of the *Trypanosoma brucei rhodesiense* metacyclic variable antigen repertoire

J. D. Barry, J. S. Crowe & K. Vickerman

Zoology Department, Glasgow University, Glasgow G12 8QQ, UK

Trypanosoma brucei rhodesiense undergoes antigenic variation in its mammalian host by changing the glycoprotein composing its surface coat^{1,2}. Trypanosome clones which have the same repertoire of variable antigen types (VATs) are said to belong to the same serodeme. Tsetse flies infected with a particular serodeme extrude infective metacyclic trypanosomes which express only a restricted part of this repertoire³⁻⁸. As the only known acquired immunity in African trypanosomiasis is VAT-specific this limitation of metacyclic VAT (M-VAT) repertoire could be important in devising a vaccine. This possibility of immunoprophylaxis could depend, however, on whether or not the M-VAT repertoire is conserved over long periods of repeated cyclical transmission and between epidemics. Studies reported here on isolates made from an East African focus of sleeping sickness over a 20-yr period suggest substantial changes in the M-VATs expressed during this time. Furthermore, we have detected change in expression of 3 M-VATs during sequential tsetse transmission of a clone in the laboratory indicating a possible instability in the organization of M-VAT genes.

Trypanosoma brucei rhodesiense, causative agent of East African sleeping sickness in man, also infects domestic and game animals. As the sole criterion for distinguishing *T. b. rhodesiense* from *T. b. brucei* is the former's infectivity to man, any *T. brucei* group stock isolated from reservoir host or tsetse fly is designated *T. brucei* ssp. Twenty-six stocks of *T. brucei* ssp. isolated from

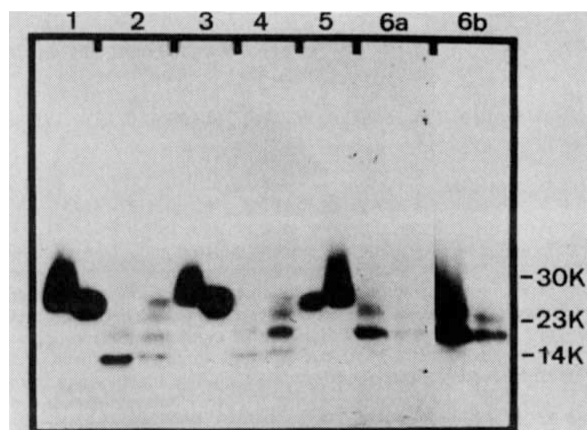


Fig. 2 Reactivity of anti-peptide 86 antibodies with thymus and brain extracts from C57BL/6J and AKR/J mice and an Nb rat. The gel was run and blotted as described in Fig. 2 legend. Panels 1 to 6b show pairs of lanes containing thymus (left) and brain (right) extracts. Panel 1, C57BL/6J (untreated); panel 2, C57BL/6J (deglycosylated); panel 3, AKR/J (untreated); panel 4, AKR/J (deglycosylated); panel 5, Nb rat (untreated); 6a and 6b, Nb rat (deglycosylated). 6a and 6b show, respectively, short and long autorad exposures of the same blot.

Table 1 Percentage GUTat variable antigen type composition of metacyclic populations of a series of trypanosome stocks

Stock	Year of isolation	7.1		7.2	7.7	7.13		7.14		7.15		7.16	7.17	5.1
		R	Mc	R	R	M	Mc	R	Mc	M	Mc	M	R	Mc
A	1961	6.5	12.5	12.0	$\frac{0}{4,080}$	14.12	12.26	1.6	ND	15.6	19.3	$\frac{0}{43,360}$	$\frac{2,200}{1,020}$	27.11
B	1961	18.97	12.0	11.08	$\frac{0}{129}$	9.12	13.20	0.5	ND	16.06	14.9	21.07	$\frac{0}{1,020}$	21.55
C	1961	14.0	12.06	6.73	$\frac{0}{4,750}$	$\frac{4,680}{810}$	$\frac{0}{810}$	0.6	1.5	17.07	18.89	$\frac{0}{2,800}$	$\frac{8,120}{8,708}$	15.70
D	1964	20.5	13.99	12.0	$\frac{0}{17,400}$	10.78	14.77	0.56	1.27	20.41	21.29	$\frac{0}{1,690}$	$\frac{8,708}{8,708}$	20.92
E	1976	21.23	19.0	13.0	ND	6.0	11.06	$\frac{0}{42,278}$	ND	15.16	14.5	7.05	ND	$\frac{0}{39,260}$
F	1977	40.13	31.0	10.57	$\pm$	12.38	9.69	$\frac{0}{64,874}$	$\frac{0}{1,300}$	$\pm$	$\pm$	17.37	$\pm$	$\frac{0}{68,540}$
G	1979	13.0	12.16	5.0	$\frac{0}{240}$	7.66	9.47	$\frac{0}{200}$	ND	9.31	15.96	44.86	$\frac{0}{672}$	$\frac{0}{38,733}$

Results of VAT-specific indirect immunofluorescence³ are expressed as percentages, calculated from a minimum of 100 but usually more than 1,000 metacyclic forms examined per sample. For negative samples, the actual number counted is shown and usually includes total probed populations from more than one fly. Tsetse flies infected with these stocks can produce several hundred to 60,000 metacyclic forms per probe. R, VAT-specific rabbit antiserum; M, VAT-specific mouse antiserum; Mc, monoclonal antibody; ND, not determined; $\pm$ = not always present in the stock (see Table 2). The monoclonal antibodies used were: GUPM 5.1, 11.2 and 17.2 (for VAT GUTat 7.1); GUPM 17.1 (GUTat 7.13); GUPM 6.1 (GUTat 7.14); GUPM 10.2 (GUTat 7.15) and GUPM 10.1 (GUTat 5.1).

man, ungulates, carnivores and tsetse flies in the sleeping sickness focus along the north-east shore of Lake Victoria between 1958 and 1979 were cryopreserved at the East African Trypanosomiasis Research Organization. We screened these stocks for common VATs using the immune lysis reaction and polyvalent sera from rabbits infected for at least 1 month with trypanosomes of the different stocks. Thirteen of the 16 stocks shown to share VATs extensively were cloned and the clones confirmed as belonging to the same serodeme. Genetic similarity was also demonstrated by analysis of electrophoretic variation in 20 enzymes (A. Tait and J.D.B., unpublished); phenotypic homogeneity was confirmed by application of several diversity indices which demonstrated greater homogeneity than *T. brucei* stocks examined previously⁹. Seven of the cloned stocks (five isolated from man) were selected for the study of their M-VATs.

Trypanosomes were grown in mice and transmitted by *Glossina morsitans* as described previously¹⁰. Eleven M-VATs were investigated by indirect immunofluorescence. M-VATs GUTat (Glasgow University *Trypanozoon* antigen type) 7.1, 7.2, 7.7, 7.14 and 7.17 were identified in tsetse salivary probes using specific rabbit antisera raised by infection with clones of bloodstream form trypanosomes expressing these M-VATs³ and GUTat 7.13, 7.15 and 7.16 were identified using similar specific antisera from mice. Monoclonal antibodies against each of GUTat 7.1, 7.13, 7.14, 7.15 and 5.1 were also used. In addition, two other M-VATs were identified using the monoclonal antibodies GUPM (Glasgow University Protozoology) 4.1 and 11.1. Monoclonal antibody derivation was as described previously⁷. As many metacyclic stages as possible were examined in each sample, but the low numbers of infected flies which are routinely obtained and in many cases the low yield of metacyclic forms per fly have constrained some of the analysis.

As expected for stocks belonging to the same serodeme there was a similarity in M-VAT composition in salivary probes (Table 1). However, only 3 of the 11 M-VATs, namely GUTat 7.1, 7.2 and 7.15, could be detected in all of the stocks. Some, GUTat 7.14 and 5.1, were present in all stocks isolated in the period 1961–64 but absent from those isolated in 1976, 1977 and 1979. When present GUTat 7.14 was a minor component (0.5%–1.6%) of the metacyclic population while GUTat 5.1 was major (15%–27%). GUTat 7.13 and 7.16 were not present in all stocks but for these VATs presence and absence did not correlate with date of stock isolation, as can be seen in the three stocks from a single epidemic in 1961. When present both these M-VATs occurred at high levels. GUTat 7.7 and 7.17 were generally absent, occurring very rarely and at low frequency (see below). The two M-VATs identified respectively by the monoclonal antibodies GUPM 4.1 and 11.1, were seen also to occur only sporadically and at very low levels (<0.1%; data not shown). Thus, the majority (9/11) of M-VATs were not

Table 2 Presence of individual metacyclic GUTat variable antigen types 7.15, 7.7 and 7.17 during sequential fly transmission of a trypanosome clone

Fly transmission	7.15	7.7	7.17
1	15.0		
2	13.57; 14.57		
3a	10.55; 11.0; 0/>1,000	0/5425	0/3,614
3b	13.2; 14.6; 22.0; 24.4; 0.126 0/1,810; 0/1,238; 0/2,640	0/990	0/4,960
4a	0/>1,000		
4b	0/218; 0/595; 0/2,870; 0/3,755; 0/4,476		
4c	10.11; 11.39; 11.64; 12.15; 14.55; 16.03; 20.09; 23.88		
5	0/220; 0/618, 0/1,656; 0/1,900; 0/3,650; 0/4,250	0/96	0/41
6	0/344, 0/>1,000, 0/>1,000	0/236	0/217
7	0/2,110; 0/4,046; 0/55,289	0/83	0/74
8	0/>1,000; 0/>1,000		
9	0/>1,000	0/52; 0/230; 0/540; 0/805; 0/1,060; 0/1,118; 0/3,265	0/51, 0/221, 0/517; 0/1,315; 0/2,285; 0/3,990
10	0/236; 0/400; 0/>1,000	0/655; 0/797; 0/1,080; 0/2,490 Fly 4104 D24 2/1,100 D48 0/6840 Fly 4145 D21 0/340 D68 1/3,160	0/170; 0/651; 0/1380; 0/6,830 Fly 4104 D24 5/273 D48 0/24,370 Fly 4145 D21 4/278 D68 0/9,600

Results of VAT-specific indirect immunofluorescence are expressed as percentages or as number of positives per total number examined. Each count represents a single probe from a different fly. For sequential transmissions, tsetse flies were fed on mice infected directly by fly bite or within three syringe passages after fly-transmitted infection. From transmission 3a a single trypanosome from the metacyclic population not containing GUTat 7.15 was cloned in a mouse and this was used to infect transmission 4a flies. Transmission 3b was initiated from a cryopreserved stablate of the population used for 3a and can thus be considered as a repeat of 3a. Transmission 4b was initiated from a mouse bitten by the fly in transmission 3b containing 0.126% GUTat 7.15. Transmission 4c was initiated from a mouse bitten by a 3a fly positive for GUTat 7.15. GUTat 7.15 was analysed by immunofluorescence with monoclonal antibody GUPM 10.2. On some of the negative probes a specific mouse antiserum was applied in immunofluorescence after attempted monoclonal antibody staining to confirm lack of reaction. GUTat 7.7 and 7.17 were analysed with specific rabbit antisera. For these two VATs results from sequential probes from two individual flies are shown, to demonstrate the sporadic nature of expression. D24 = day 24 after uptake of the initial infective feed by the tsetse fly.

expressed consistently in the metacyclic population (see also below).

To investigate the stability of the M-VAT repertoire in more controlled conditions, a clone of stock F was transmitted through tsetse flies 10 times sequentially, its M-VATs being examined at each transmission. Most of the M-VATs were expressed consistently during all transmissions, but three special cases were observed (Table 2). GUTat 7.15 began to disappear from the metacyclic population at the third transmission. This was noticed first by its total absence from some flies while others retained expression at the usual level (10–24%) although in one fly the VAT was expressed at a much reduced level (0.126%). A metacyclic stage trypanosome from a negative fly was cloned and subjected to a further seven transmissions, in none of which was GUTat 7.15 ever found; the loss seems to have been irreversible. Similarly, further fly transmission (4b, Table 2) of trypanosomes from the fly containing 0.126% of GUTat 7.15 resulted in loss of that VAT from the metacyclic population. In contrast, one further transmission (4c) of trypanosomes from a fly containing normal levels of GUTat 7.15 showed retention of the VAT at similar levels in all flies. It would seem that instability for expression of GUTat 7.15 arose in part of the trypanosome population following transmission 2. At the tenth sequential transmission two novel M-VATs, GUTat 7.7 and 7.17, appeared in a minority of flies, at a low level and only sporadically in individual flies (Table 2). Further transmissions will reveal whether or not they become major M-VATs.

Loss of expression of a VAT in the metacyclic population does not necessarily mean that its expression is no longer possible. Thus, GUTat 7.14 is still expressed early by bloodstream form trypanosomes in mammals bitten by tsetse flies infected with stocks E, F and G. Antibody neutralization and cloning experiments have indicated that this is due to *de novo* expression of VAT following tsetse fly bite rather than presence *ab initio* in the metacyclic population. On the other hand, it has not proved possible to detect expression of GUTat 5.1 in bloodstream infections with those stocks where it is not present as an M-VAT; perhaps the gene has been deleted or is no longer under expression control.

The metacyclic population of *T. brucei* is thus seen to undergo a drift in antigen expression, and very few VATs were not affected in the relatively short time scale within which our material was collected. While the continuous gene switching characteristic of the ontogeny of antigenic variation is an expedient system for maintenance of infection in the face of an immune response, there is a constant longer term selection for diversification of VAT repertoires and generation of new VATs. It would seem that one limitation on the ontogeny of antigenic variation is the predictability of M-VAT expression within a single stock; repeated challenge of the symptomless reservoir hosts with the same M-VATs could lead to immunity. Our results suggest that trypanosome populations in the field effectively overcome this limitation by a gradual change in VAT composition of the metacyclic population.

Major change in the genomic context of VAT genes independent of expression and involving translocation, conversion and even total deletion have been described elsewhere^{11–13}. It would seem that such rearrangements play an important role in the evolution of both the expression sequence and the gene content of the VAT repertoire. Interestingly, the genes most prone to these changes appear to be telomeric; preliminary studies (unpublished) on two of the M-VATs of this serodeme suggest that both are telomeric. Preferential expression in the salivary gland population is possibly imposed by the genetic mechanisms of antigenic variation, the M-VAT genes having a positional or recognition advantage within the genome. That some VATs can be lost from the metacyclic population but retain early expression in the mammal while others may be lost entirely could be due to differential expression of different members of gene families in the metacyclic and bloodstream populations or could involve a variety of genomic rearrangements, some affecting only metacyclic expression and others total expression. The

dynamic trypanosome genome continues to reveal intriguing features.

This work was supported by the MRC and the UNDP/World Bank/WHO Special Programme for Research and Training on Tropical Diseases. We thank Dr A. Jordan, Tsetse Research Laboratory, Bristol, for supplying tsetse fly pupae and Ms C. R. Cameron for technical assistance.

Received 22 August; accepted 4 October 1983.

1. Vickerman, K. & Barry, J. D. in *Immunology of Parasitic Infections* (eds Cohen, S. & Warren, K. S.) 204–260 (Blackwell, Oxford, 1982).
2. Vickerman, K. *Nature* **273**, 613–617 (1978).
3. Barry, J. D., Hajduk, S. L., Vickerman, K. & Le Ray, D. *Trans. R. Soc. trop. Med. Hyg.* **73**, 205–208 (1979).
4. Barry, J. D. & Hajduk, S. L. in *Pathogenicity of Trypanosomes* (eds Losos, G. & Chouinard, A.) 51–56 (IDRC, Ottawa, Canada, 1979).
5. Hajduk, S. L., Cameron, C. R., Barry, J. D. & Vickerman, K. *Parasitology* **83**, 595–607 (1981).
6. Hajduk, S. L. & Vickerman, K. *Parasitology* **83**, 609–621 (1981).
7. Crowe, J. S., Barry, J. D., Hajduk, S. L. & Vickerman, K. *Trans. R. Soc. trop. Med. Hyg.* **75**, 894–895 (1981).
8. Crowe, J. S., Barry, J. D., Luckins, A. G., Ross, C. A. & Vickerman, K. *Nature* **306**, 389–391 (1983).
9. Tait, A. *Nature* **287**, 536–538 (1980).
10. Le Ray, D., Barry, J. D., Easton, C. & Vickerman, K. *Ann. Soc. belge Med. trop.* **57**, 369–381 (1977).
11. Borst, P. et al. in *Proc. Cetus-UCLA Meeting on Molecular Biology of Host-Parasite Interactions* (eds Agabian, N., Eisen, H.) (Park City, Utah, in the press).
12. Borst, P. et al. *UCLA Symp. molec. cell. Biol. New Ser.* **8**, (in the press).
13. Young, J. R., Shah, J. S., Matthysens, G. & Williams, R. O. *Cell* **32**, 1149–1159 (1983).

Close linkage of fragile X-mental retardation syndrome to haemophilia B and transmission through a normal male

G. Camerino*, M. G. Mattei†, J. F. Mattei†, M. Jaye‡ & J. L. Mandel*

* Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS, Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM, 11, rue Humann, 67085 Strasbourg, France

† Centre de Génétique Médicale, INSERM U242, Hôpital de la Timone, 13385 Marseille, France

‡ TRANSGENE, 11, rue Humann, 67000 Strasbourg, France

The fragile X-mental retardation syndrome¹ is defined by a moderate to severe mental retardation associated with a cytogenetic marker, a fragile site localized on the long arm of the X chromosome at band Xq 27 (ref. 2). This syndrome has recently been recognized as one of the major causes of genetically determined mental retardation, and as one of the most important X-linked diseases with respect to its frequency (analogous to that of Duchenne muscular dystrophy or of haemophilia A) and severity^{3–6}. In the absence of treatment, genetic screening for this disease would seem particularly important. Prenatal diagnosis is now feasible although difficult^{7,8} and detection of heterozygous carriers is only possible in ~50% of cases^{1,4,9}. The recent demonstration of genetic linkage between the glucose 6-phosphate dehydrogenase (G6PD)-colour blindness cluster (at Xq28) and the fragile X locus has suggested that the fragile site is indeed the site of the mutation¹⁰. We show here that the fragile X and haemophilia B loci are closely linked, using as genetic marker a polymorphism of the coagulation factor IX gene¹¹. Our study of a large family has demonstrated transmission through a phenotypically normal male, a feature previously described in retrospective analysis of a few other fragile X pedigrees^{1,6,12,13}. Restriction polymorphisms associated with the factor IX gene should be useful for analysing this peculiar aspect of the genetics of the fragile X syndrome, and for genetic screening of the disease.

Using a cloned cDNA corresponding to human coagulation factor IX¹⁴, we have assigned the gene (whose malfunction is responsible for haemophilia B) to the q26–q27 region of the human X chromosome¹¹ (however, localization in the proximal

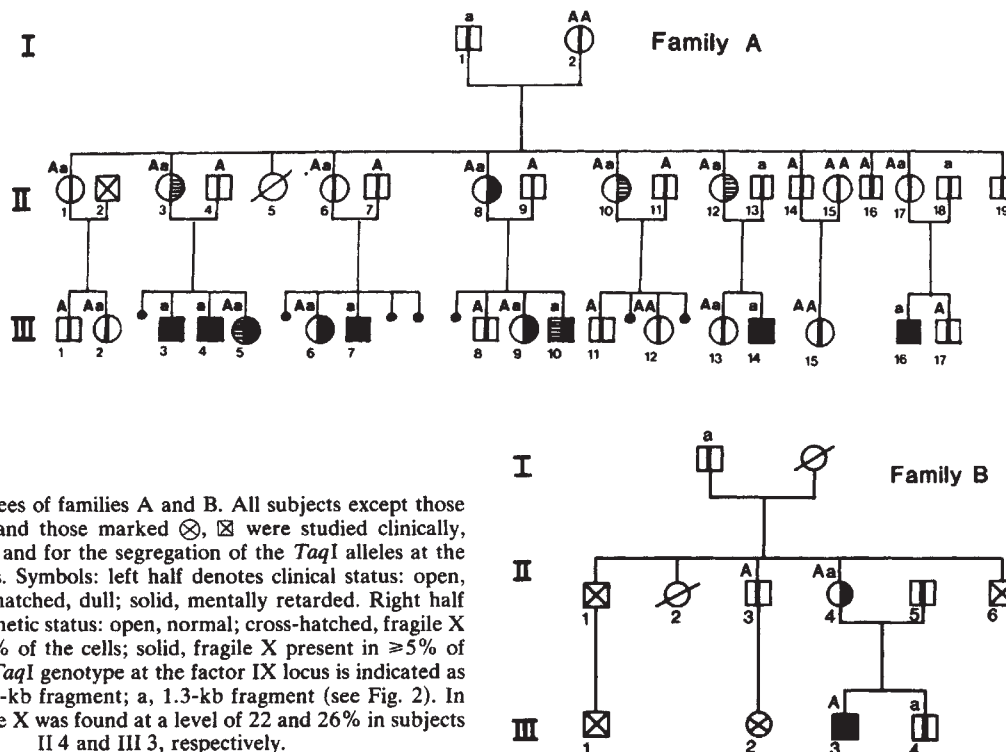


Fig. 1 Pedigrees of families A and B. All subjects except those deceased (⊗) and those marked ⊗, ⊗ were studied clinically, cytogenetically and for the segregation of the *TaqI* alleles at the factor IX locus. Symbols: left half denotes clinical status: open, normal; cross-hatched, dull; solid, mentally retarded. Right half denotes cytogenetic status: open, normal; cross-hatched, fragile X present in $\leq 2\%$ of the cells; solid, fragile X present in $\geq 5\%$ of the cells. The *TaqI* genotype at the factor IX locus is indicated as follows: A, 1.8-kb fragment; a, 1.3-kb fragment (see Fig. 2). In family B, fragile X was found at a level of 22 and 26% in subjects II 4 and III 3, respectively.

Table 1 Expression of the fragile X and segregation of factor IX alleles in members of family A

Pedigree no.	Sex	Age (yr)	Mental status	Fragile X cells / lymphocytes	% Fragile X	Factor IX alleles
I 1	M	59	N	0/200	0	a
I 2	F	56	N	0/200	0	AA
II 1	F	36	N	0/100	0	AA
II 3	F	35	N	2/130	1.5	Aa
II 6	F	32	N	0/100	0	Aa
II 8	F	28	N	3/60	5.0	Aa
II 10	F	27	N	2/100	2.0	Aa
II 12	F	25	N	1/100	1.0	Aa
II 14	M	24	N	0/50	0	A
II 16	M	22	N	0/50	0	A
II 17	F	19	N	0/100	0	Aa
II 19	M	17	N	0/50	0	(-)
III 1	M	14	N	0/50	0	A
III 2	F	12	N	0/50	0	Aa
III 3	M	16	MR	18/50	36.0	a
III 4	M	14	MR	28/50	56.0	a
III 5	F	12	Dull	17/50	34.0	Aa
III 6	F	7	N	8/50	16.0	Aa
III 7	M	6	MR	6/50	12.0	a
III 8	M	7	N	0/100	0	A
III 9	F	4	N	20/50	40.0	Aa
III 10	M	1	Subnormal	3/50	6.0	a
III 11	M	5	N	0/100	0	A
III 12	F	2	N	0/50	0	AA
III 13	F	4	N	0/50	0	Aa
III 14	M	3	MR	20/50	40.0	a
III 16	M	2	MR	26/50	52.0	a
III 17	M	1	N	0/60	0	A

part of q28 cannot be excluded). Furthermore, analysis of a hybrid cell line containing a fragment of the long arm of the human X chromosome, with probes for factor IX, hypoxanthine phosphoribosyltransferase (HPRT) and sequences from the q24-q26 region, suggested that the factor IX gene is distal with respect to the HPRT gene (unpublished results). The factor IX cDNA probe detects a *TaqI* restriction fragment length polymorphism in the 5' region of the gene, with a frequency that is useful for genetic analysis, as about 40% of females are heterozygous for this marker¹¹. The localization of factor IX gene suggested that it might show measurable linkage with the fragile X-mental retardation locus, as the fragile site is close to the q27-q28 interface². A first screening yielded two families informative for such linkage analysis. Family A (Fig. 1) is a large three-generation family, particularly favourable for genetic analysis. Five of the seven sisters in generation II have affected sons and are thus obligatory carriers, although the cytogenetic analysis (Table 1) yields results which are either negative (for II 6 and 17) or at the limit of significance (1-2% of fragile X in II 3 and 12). The carrier status of their daughters can be assigned unambiguously (Table 1) due to the known inverse relationship between age and frequency of the fragile site in heterozygous females^{9,15}. All seven sisters in generation II are heterozygous for the restriction polymorphism detected by the factor IX probe (Aa genotype; Figs 1, 2).

When the progeny of the five obligatory carriers was analysed, a striking result was observed. All nine affected sons or carrier daughters (III 3, 4, 5, 6, 7, 9, 10, 14, 16) had inherited the 1.3-kilobase (kb) *TaqI* fragment (a allele) from their mother, while the two normal sons (III 8 and 17) and one non-carrier daughter (III 13) had inherited the 1.8-kb fragment (A allele) from their mother (Fig. 1, Table 1). If one assumes that the fragile site was transmitted by the grandmother (I 2, homozygous AA) as expected for an X-linked trait, this would mean that there were 12 recombination events in the 12 informative meioses analysed. Even if the two loci were unlinked, the chance of observing such a result would be very small (0.5^{12} or 1 in 4,096).

The alternative hypothesis would be that the fragile X was transmitted by the phenotypically normal grandfather who carries the *TaqI* a allele, and in this case there would be no

Presence of the fragile site on the X chromosome was analysed in lymphocytes as described previously³². Only those chromosomes showing clear displacement of the terminal band were scored as positive. For generation III the factor IX allele which is underlined is that transmitted by the mother. Mental status is described as normal (N), dull or mentally retarded (MR). Expression of the mental retardation cannot be fully assessed in children as young as III 10. All other members of family A which were studied are clinically and cytogenetically normal and their genotype at the factor IX locus is indicated in Fig. 1.

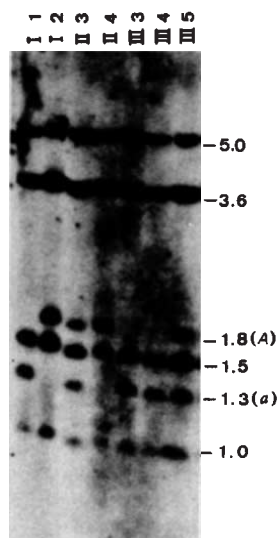


Fig. 2 Southern blot analysis of *TaqI*-digested DNAs from members of family A. The DNAs were prepared, digested with *TaqI*, transferred to diazobenzyloxymethyl paper and hybridized as described previously^{11,33}. The factor IX cDNA probe used corresponds to the 5' half of the cloned sequence¹⁴ and does not hybridize to the 2.7-kb *TaqI* fragment of the gene¹¹. Numbering of lanes corresponds to pedigree numbers in Fig. 1. The sizes of the DNA fragments are given in kilobases.

recombinants in the 12 meioses. With the latter hypothesis, all brothers in generation II should be normal and all sisters should be carriers. The first condition is indeed observed for the three brothers, however the carrier status of the females can only be ascertained through their progeny, as even the obligatory carriers in this family do not always show the fragile site at a significant level. It can be calculated, however, that the probability of obtaining such a family (three normal males, five out of seven sisters transmitting the trait to their progeny of two or three children) is 36 times higher under the hypothesis of transmission through the normal grandfather than under the more classical hypothesis of carrier female (I2) transmission. This value is independent of the probability calculated from the recombination data, and the combined likelihood for male transmission (with close linkage between the loci analysed) is 147,000 times ($36 \times 4,096$) that of female transmission (with no linkage), which validates the former hypothesis. (There was no indication of other fragile X cases in the families of the grandfather or grandmother.) Thus, daughters II 1 and II 10 must be fragile X carriers (in fact II 10 shows 2% of fragile X, a value indicative of heterozygosity), and their normal progeny can be analysed for the restriction site marker. In the three informative cases (III 1, 11, 12) the A allele was found to be inherited from the mother (a conclusion cannot be reached for III 2 as the father was unavailable for study).

From this single family, it can be concluded that there were no recombinants out of 15 informative meioses. In the second family analysed (Fig. 1), the normal child had the same 'a' allele as the normal grandfather, while the affected one had the 'A' allele. Thus, assuming 'normal' transmission from the grandmother, there is no recombination in the two informative meioses (the alternative hypothesis of male transmission would imply two recombination events, which is unlikely in view of the close linkage already indicated by family A).

The pooled results (17 non-recombinants, 0 recombinants) provide a relative probability for linkage of 130,000 at a recombination value of 0% (in the classical logarithmic form this corresponds to a lod (log of odds) score¹⁶ of 5.12). From the relative probability curve (Fig. 3) it can be estimated that the distance between the two loci is <12 centimorgans at a 90%

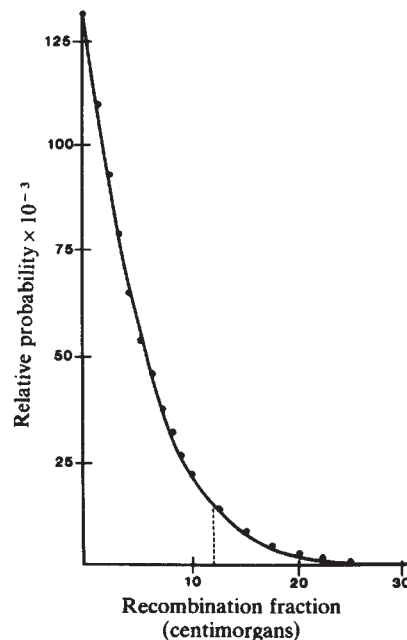


Fig. 3 Linkage between the fragile X locus and the factor IX gene. The relative probability of linkage¹⁶ is plotted for various values of the recombination fraction θ . The peak at $\theta = 0$ corresponds to a lod score (log of odds) of 5.12. The broken line indicates the upper limit for the genetic distance between the two loci, at a 90% probability¹⁶.

confidence level. Thus, close linkage of the fragile X syndrome with the haemophilia B locus is more firmly established than linkage with the G6PD-colour blindness cluster¹⁰ as in the latter case the maximum likelihood estimate for recombination was 6% (6 centimorgans) with a lod score of 3.16 and wide 90% confidence limits. One would expect that the haemophilia B locus would show linkage with the G6PD cluster at Xq28, in apparent contradiction with a published study¹⁷. This suggests that the factor IX and G6PD loci flank the fragile site, and might be more than 20 centimorgans apart. The high frequency of the restriction polymorphism associated with the factor IX gene¹¹ should allow us to assay for linkage to the G6PD cluster on the distal side and to the HPRT locus¹⁸ on the proximal one and to refine the genetic map of the X chromosome in this area.

The present findings are significant in two respects. The factor IX-associated polymorphism should provide a tool for studying the puzzling problem of transmission through normal males of the fragile X trait, and also seems promising for use in genetic screening of this disease. A few family cases of male transmission of the fragile X syndrome have already been described^{1,6,12,13}, based, with two exceptions^{6,19}, on retrospective analysis of large pedigrees, the putative male transmitters being dead. In the case of family A, the grandfather who transmitted the trait shows neither the cytogenetic marker nor the characteristic clinical features of the disease. The occurrence of such families poses the fundamental problem of explaining such a segregation, and may also lead to errors in genetic counselling (as the trait would appear in unexpected branches of the families) and in linkage studies in fragile X families. Because of the paucity of large families, one generally has to pool linkage data from several small families, where male transmission might not be recognized. In such cases, recombinants could be mistaken for non-recombinants and vice versa. Closely linked markers such as that used here could allow us to recognize and estimate the frequency of male transmission, even in relatively small families.

Genetic screening based on analysis at the DNA level is already available for a few genetic defects where the corresponding gene has been cloned²⁰⁻²³. This has several advantages over conventional methods as it is independent of the site of expression of the gene, and prospects exist for using chorionic

villi biopsies²⁴ to allow early prenatal diagnosis. It has been proposed that even when the genes have not been isolated or are not defined biochemically, linked restriction polymorphisms could be used to screen for diseases such as Duchenne muscular dystrophy (DMD), Huntington's chorea or cystic fibrosis^{25,26}. Polymorphic probes showing linkage to DMD have recently been described^{26,27} and have proved important in confirming the localization of the Duchenne locus in the Xp21 region, and for analysing X-linked myopathies having variant clinical features²⁸. Their diagnostic value is limited, however, as they are only loosely linked to the disease locus (≈ 17 centimorgans). Other potentially useful linkages are those of G6PD with haemophilia A^{17,29}, adenoleukodystrophy³⁰ and the fragile X syndrome¹⁰. Especially in the latter case, the distance between the marker and the disease locus needs to be estimated more precisely. Furthermore, no polymorphism has yet been reported for the cloned G6PD sequences³¹ (at the protein level, this marker is only polymorphic in certain populations). At present, the factor IX marker seems to offer the best prospects for use in screening of the fragile X syndrome, and we have already demonstrated in two cases a carrier status which had been overlooked on the basis of the cytogenetic analysis. More studies are needed to refine the linkage data and to find additional polymorphisms associated with the factor IX probe. In conjunction with an informative marker on the other side of the fragile X locus, this would provide a reliable and generally usable screening test.

We thank Dr M. Auger (Marseille) for invaluable help in the family study, the Institut Merieux (Lyon) for allowing us to use the factor IX probe, Dr P. Tolstoshev for editorial help and Mrs C. Kutschis and C. Werlé for preparing the manuscript. This work was supported by grants from INSERM (PRC 134029 and CRL 826032) and by an EMBO long-term fellowship to G.C.

Received 22 August; accepted 19 October 1983.

1. Tariverdian, G. & Weck, B. *Hum. Genet.* **62**, 95-109 (1982).
2. Brookwell, R. & Turner, G. *Hum. Genet.* **63**, 77 (1983).
3. Herbst, D. S. & Miller, J. R. *Am. J. med. Genet.* **7**, 461-469 (1980).
4. Mattei, J. F., Mattei, M. G., Aumeras, C., Auger, M. & Giraud, F. *Hum. Genet.* **59**, 281-289 (1981).
5. Fishburn, J., Turner, G., Daniel, A. & Brookwell, R. *Am. J. med. Genet.* **14**, 713-724 (1983).
6. Jacobs, P. A., Mayer, M., Matsuura, J., Rhoads, F. & Yee, S. C. *Hum. Genet.* **63**, 139-148 (1983).
7. Shapiro, L. R. *et al. Lancet* **i**, 99 (1982).
8. Sutherland, G. R. & Jacky, P. B. *Lancet* **i**, 100 (1982).
9. Sutherland, G. R. *Hum. Genet.* **53**, 23-27 (1979).
10. Filippi, G. *et al. Am. J. med. Genet.* **15**, 113-119 (1983).
11. Camerino, G. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
12. Nielsen, K. B., Tommerup, N., Poulsen, H. & Mikkelsen, M. *Hum. Genet.* **59**, 23-25 (1981).
13. Fryns, J. P. & Van Den Berghe, H. *Hum. Genet.* **61**, 262-263 (1982).
14. Jaye, M. *et al. Nucleic Acids Res.* **11**, 2325-2335 (1983).
15. Chudley, A. E. *et al. Am. J. med. Genet.* **14**, 699-712 (1983).
16. Emery, A. E. H. in *Methods in Medical Genetics*, 63-75 (Churchill-Livingstone, Edinburgh, 1976).
17. Whittaker, D. L., Copeland, D. L. & Graham, J. B. *Am. J. hum. Genet.* **14**, 149-158 (1962).
18. Nussbaum, R. L., Crowder, W. E., Nyhan, W. L. & Caskey, T. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4035-4039 (1983).
19. Webb, G. C., Rogers, J. G., Pitt, D. B., Halliday, J. & Theobald, T. *Lancet* **ii**, 1231-1232 (1981).
20. Little, P. F. R. in *Genetic Engineering* Vol. 1 (ed. Williamson, R.) (Academic, New York, 1981).
21. Wilson, J. T. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 3628-3631 (1982).
22. Conner, B. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 278-282 (1983).
23. Kidd, V. J., Wallace, R. B., Itakura, K. & Woo, S. L. C. *Nature* **304**, 230-234 (1983).
24. Elles, R. G., Williamson, R., Niaz, M., Coleman, D. V. & Horwell, D. *New Engl. J. Med.* **308**, 1433-1435 (1983).
25. Botstein, D., White, R. L., Scolnick, M. & Davis, R. W. *Am. J. hum. Genet.* **32**, 314-331 (1980).
26. Murray, J. M. *et al. Nature* **300**, 69-71 (1982).
27. Davies, K. E. *et al. Nucleic Acids Res.* **11**, 2303-2312 (1983).
28. Kingston, H. M., Thomas, N. S. T., Pearson, P. L., Sarfarazi, M. & Harper, P. S. *J. med. Genet.* **20**, 255-258 (1983).
29. Tsevenis, H., Mandalaki, T., Volkens, W. S. & Khan, P. M. *Cytogenet. Cell Genet.* **25**, 213 (1979).
30. Migeon, B. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 5066-5070 (1981).
31. Persico, M. G., Toniolo, D., Nobile, C., D'Urso, M. & Luzzato, L. *Nature* **294**, 778-780 (1981).
32. Mattei, M. G., Mattei, J. F., Vidal, I. & Giraud, F. *Hum. Genet.* **59**, 166-169 (1981).
33. Hanauer, A. *et al. Nucleic Acids Res.* **11**, 3503-3516 (1983).

A yeast gene encoding a protein homologous to the human *c-has/bas* proto-oncogene product

Dieter Gallwitz*, Cornelia Donath* & Christian Sander†

* Institut für Physiologische Chemie I, Universität Marburg, D-3550 Marburg/Lahn, FRG

† Abteilung Biophysik, Max-Planck-Institut für Medizinische Forschung, D-6900 Heidelberg, FRG

Organisms amenable to easy genetic analysis should prove helpful in assessing the function of at least those proto-oncogene products which are highly conserved in different eukaryotic cells. One obvious possibility is to pursue the matter in *Drosophila melanogaster* DNA, which has sequences homologous to several vertebrate oncogenes¹⁻³. Another is to turn to the yeast *Saccharomyces cerevisiae*, if it contains proto-oncogene sequences. Here we report the identification of a gene in *S. cerevisiae* which codes for a 206 amino acid protein (YP2) that exhibits striking homology to the p21 products of the human *c-has/bas* proto-oncogenes^{4,5} and the transforming p21 proteins of the Harvey (*v-ras*^H)⁶ and Kirsten (*v-ras*^K)⁷ murine sarcoma viral oncogenes. The YP2 gene is located between the actin^{8,9} and the tubulin gene¹⁰ on chromosome VI¹¹ and is expressed in growing cells. The protein it encodes might share the nucleotide-binding capacity of p21 proteins²²⁻²⁴.

Previously we have isolated two recombinant plasmids (pYA102 and pYA208) containing overlapping DNA fragments which harbour the yeast actin gene^{8,9}. In studying the region 5' to the actin gene we detected another gene whose transcription product is a polysomal RNA coding for a protein of $\sim 20,000$ molecular weight (MW)⁸. This protein, provisionally termed 'protein 2', will be subsequently referred to as YP2 protein.

The location of the YP2 coding region was mapped with neighbouring DNA fragments which were ³²P-nick-translated and used to identify the YP2 mRNA on RNA blots. A 2.18-kilobase (kb) *HindIII*-*PstI* fragment spanning the region from the actin gene codon 260 to nucleotide 960 upstream from the actin mRNA cap site^{9,12} (fragment A in Fig. 1) and a 0.79-kb *XhoI*-*HincII* fragment originating in the actin gene intron and ending 577 base pairs (bp) 5' to the actin mRNA start site (fragment C in Fig. 1), hybridized to the 1,430-nucleotide actin mRNA and to the 830-nucleotide YP2 mRNA (as shown in Fig. 2). A 1.72-kb *HindIII*-*PstI* fragment with the *HindIII* restriction site residing in codon 391 of the β -tubulin gene¹⁰ hybridized, in addition to the tubulin gene transcript of about 1,600 nucleotides, also to the YP2 mRNA (Fig. 2). These results established that the unique *PstI* and *HincII* restriction sites must reside within the YP2 gene.

The nucleotide sequence of the YP2 gene region was determined by the chemical sequencing method of Maxam and Gilbert¹³ and is presented in Fig. 3. Screening for possible reading frames on both strands led to the identification of only one open frame for a protein of about 20,000 MW. The second longest continuous stretch not interrupted by stop codons could code for a protein of only 104 amino acids. Because we found no indication for the presence of an intervening sequence within the sequenced region, neither by searching for an unspliced transcript in the temperature-sensitive splice-defective yeast *rna-2* mutant^{14,15} nor by the existence of a TACTAACA sequence which is a common structural feature of introns in all yeast polymerase II-transcribed genes¹⁶, these results established that the protein YP2 is 206 amino acid residues long (including the initiating methionine) and has a MW of 23,213.

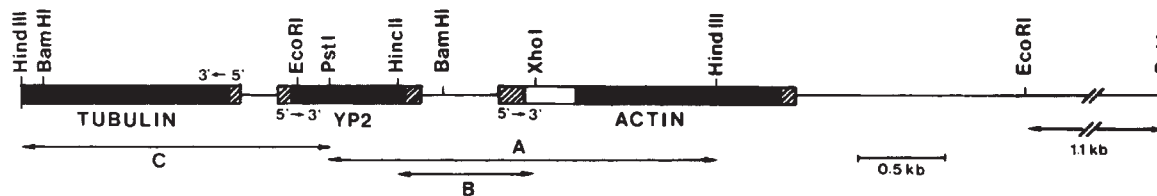


Fig. 1 Location of the YP2 gene between the actin and the β -tubulin gene on two overlapping DNA fragments cloned in the recombinant plasmids pYA102 and pYA208⁸. The direction of transcription of the three genes is shown by small arrows. The fragments (A, B and C) used to locate the YP2 gene (see Fig. 2) are indicated. Only those restriction endonuclease cutting sites mentioned in the text are shown. Solid blocks, exon; open blocks, intron; hatched blocks, untranslated regions.

Visual inspection of the YP2 amino acid sequence revealed a striking homology with the p21 proteins encoded by the human *c-has/bas* proto-oncogene^{4,5} and their viral relatives, *v-ras*^H (ref. 6) and *v-ras*^K (ref. 7). A computer-assisted comparison of the 206-residue YP2 and the 189-residue p21 proteins established that a homology of 38% exists between residues 9–171 of the YP2 and residues 4–165 of the human p21 protein with only three one-residue gaps. The homology was statistically highly significant by alignment matrix techniques^{17,18}. Sequence alignment showed (Fig. 5) four blocks of five to eight residues which match perfectly. In addition, many changes between the two sequences are conservative with respect to the chemical properties of the amino acid side chains and to the observed evolutionary amino acid replacements. The striking homology does not extend much further than the glutamine residue in positions 171 and 165 of the YP2 and the p21 protein, respectively. Interestingly, it is this residue which separates the nearly identical 165-residue segment of the transforming p21 proteins encoded by *v-ras*^H (ref. 6) and *v-ras*^K (ref. 7) viral oncogenes from their more divergent carboxyl terminal region of 24 amino acids.

The amino acid sequence homology near positions where single point mutations are responsible for the transforming ability of altered *c-ras* genes^{6,7,19,20} is particularly strong. Gly 12 of the normal human *c-has/bas* proto-oncogene product corresponds to Ser 17 of YP2; p21 *v-ras*^K also has Ser, *v-ras*^H has Arg and the activated *c-has/bas* gene products of T24 and EJ bladder carcinoma have Val in this position. Gln 61 of the normal *c-has/bas* proto-oncogene-encoded protein is preserved in YP2 as Gln 67 where Leu 61 is responsible for the

activation of the Hs242 oncogene in a human lung carcinoma-derived cell line²¹. Thr 59 which is phosphorylated in p21 *v-ras*^H and *v-ras*^K (refs 22–24) is Ala 65 in YP2, similar to Ala 59 of the normal *c-has/bas* product.

The 5' and 3' ends of the YP2 mRNA were mapped by S₁ nuclease protection experiments^{25,26} (data not shown) and it was found that there are two major cap sites, 34 and 84 nucleotides upstream from the AUG initiation codon, and seven termination sites all of which are A residues (as indicated in Fig. 3). The termination sites are preceded by three putative polyadenylation signal sequences, a situation reminiscent of the *S. cerevisiae* actin gene¹².

The YP2 gene has the same orientation as the actin gene whereas the β -tubulin gene, which was recently sequenced by Neff *et al.*¹⁰, is transcribed in the opposite direction to the YP2 gene. According to our S₁ nuclease mapping experiments, the tubulin mRNA, like the YP2 mRNA, has two major cap sites and only 195 bp separate the most distant 5' start sites of the two mRNAs (Fig. 3). An interesting feature of the YP2–tubulin

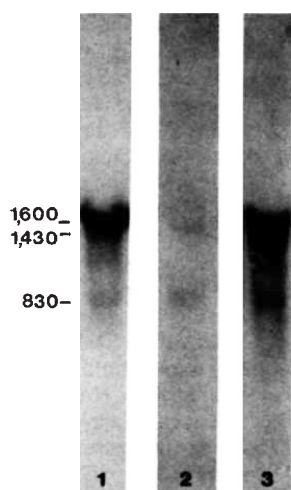


Fig. 2 Mapping of the YP2 gene through the identification of YP2 mRNA on Northern blots. The DNA fragments A, B and C (shown in Fig. 1) were nick-translated with [α -³²P]dCTP and hybridized to total cellular RNA after it had been glyoxylated, separated on 1.5% agarose gels and transferred to nitrocellulose filters as described previously¹⁶. Hybridization was to fragment C (lane 1), fragment B (lane 2) and fragment A (lane 3). The lengths (in nucleotides) of the hybridizing RNAs are indicated on the left margin.

TUBULIN
 0001 .5CGAGATGATGAATGATTCTCTCATTATATTACTTTGTGGAGATTTTGTCTTTGTAGTTAGTAGCTGCTATG
 SerIleHisIleIleGluArgMet
 0076 TCACCTCAATGGCTCACCACACTCTTCCACTAGATCTATTAAAGTTTGGTGTGTGTGTGTGTGTGTGTGT
 CAP1 CAP2
 0151 TTCTGTTTTTCTCTTTCTTGTGGCGTCGCCAGGTCACTGTACAGCTATATCTGTACCCGGCGCCAAATTC
 CAP2
 0226 CTGATAGTACTATTTTTTGGCCCTGGGGGAATAAGAAAAAGAGAAGAGCAAGCAAGCAAAATTCAGCGATTAA
 CAP1
 0301 AGTGCACCAGTTTGTAGGAGGATATCTTTACTGTTAGGCCAGGTGAAATAAATAAACAAGAAAGCACTAAT
 10 20
 MetAsnSerGluTyrAspTyrLeuPheLysLeuLeuLeuIleGlyAsnSerGlyValGlyLysSerCysLeuLeu
 0376 ATGAATAGCGAGTACGATTACCTGTTCAACTGCTGTGTGATCGGGAATTCGGGTGTCGGGAAGTCTGTGTACTT
 30 40
 LeuArgPheSerAspAspThrTyrThrAsnAspTyrIleSerThrIleGlyValAspPheLysIleLysThrVal
 0451 TTGAGTTTTCGGAGCACATATACCAACGACTACATCTCCACATTTGGAGTGAGTCAAGATTAAAGACTGTA
 60 70
 GluLeuAspGlyLysThrValLysLeuGlnIleTrpAspThrAlaGlyGlnGluArgPheArgThrIleThrSer
 0526 GAACGTGACGGCAAGACTGTAAGCTACAGATTGGGACACTGACAGTCAAGCAAGCTTTCCGACTATCACTTCA
 80 90
 SerTyrTyrArgGlySerHisGlyIleIleIleValTyrAspValThrAspGlnGluSerPheAsnGlyValLys
 0601 TCTTACTCCGTGGTTCGCATGGGATCATTATCGTGTATGATGTCACGACCAAGAAATCTTCAACGGCGTGAG
 110 120
 MetTrpLeuGlnGluIleAspArgTyrAlaThrSerThrValLeuLysLeuLeuValGlyAsnLysCysAspLeu
 0676 ATGTGGCTGACAGAGATTGATCGGTATGCAACCTCAACAGTGTGAAGCTATTGGTAGTAAACAAAGTGATTTA
 130 140
 LysAspLysArgValValGluTyrAspValAlaLysGluPheAlaAspAlaAsnLysMetProPheLeuGluThr
 0751 AAGGACAGCGTGTCTGGAAATGACGTGGCAAGGAATTCGGGACGGCAATAAGATGCGCTTTAGAGACT
 160 170
 SerAlaLeuAspSerThrAsnValGluAspAlaPheLeuThrMetAlaArgGlnIleLysGlnSerMetSerGln
 0826 AGTGCCTTTGGACTCCCAACGTCGAGGATGCGTTTTTAACCATGGCTAGACAAATCAAGCAAGATGTGCCAA
 180 190
 GlnAsnLeuAsnGluThrThrGlnLysLysGluAspLysGlyAsnValAsnLeuLysGlyGlnSerLeuThrAsn
 0901 CAAACCTGAACCAACCACTCAGAGAAGGAAGACAAAGGACGTCACCTCAAGGACAGAGTTTAAACCAAC
 206
 ThrGlyGlyGlyCysSTP
 0976 ACCGTTGGGGCTGCTGTTGAACAAGCGCGCTCTACCTTGCAGACCATATAATATAATACTAAATAGTAA
 1051 TAAGACACACGGAGAACATATATACACATTACAGTAACATAACAGAGACAGATACTACCAAAATGTGTGG
 1100
 1126 GGAAGCGGGTAAAGTGCACAGCAATTAATGCACACATTAACTACATCTTCTCTATCGGATCC 3'

Fig. 3 Nucleotide sequence of the YP2 gene and the YP2–tubulin inter-gene region. The two major cap sites of the YP2 and the tubulin mRNAs are indicated by arrowheads. Note the opposite orientation of the YP2 and tubulin gene. The multiple termination sites of the YP2 mRNA are indicated with asterisks; the presumptive polyadenylation signal sequences are underlined. The inverted repeat is shown by arrows.

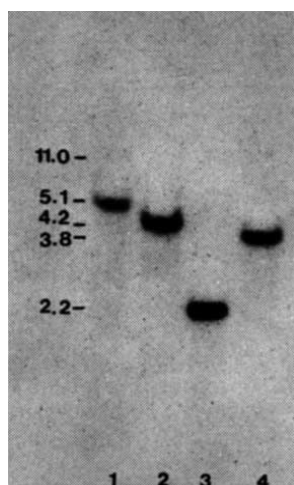


Fig. 4 Southern blot analysis to determine the YP2 gene copy number. DNA from the haploid yeast strain A364A¹⁴, which was also used to isolate the YP2 gene⁸, was digested to completion with the restriction endonucleases *Pst*I (lane 1), *Eco*RI (lane 2), *Bam*HI (lane 3) or *Hind*III (lane 4) and 10 µg of DNA were separated on a 0.9% agarose gel in 89 mM Tris-borate, 1 mM EDTA, pH 8.3 (ref. 31). The fragments were transferred bidirectionally to nitrocellulose filters³². Hybridization was performed at 68 °C in 3 × SSC, 50 mM sodium phosphate pH 6.8, 2 mM EDTA, 2 × Denhardt's solution, 50 µg ml⁻¹ sonicated, denatured calf thymus DNA and 4 × 10⁵ c.p.m. ml⁻¹ of denatured, ³²P-nick-translated 526 bp *Eco*RI/*Hinc*II fragment (specific activity 10⁸ d.p.m. µg⁻¹) from the YP2-coding region (see Fig. 1). Filters washed at 65 °C and either 1 × SSC or 0.1 × SSC gave essentially the same labelled fragments. The lengths (in kb) of the hybridizing fragments are shown on the left margin.

intergene region is an almost perfect inverted repeat of 13 nucleotides within the centre of this 195-bp segment (Fig. 3). The repeat could be folded into an energetically stable stem and loop structure ($\Delta G = -10.6$ kcal mol⁻¹)²⁷ and it is tempting to speculate that this hairpin structure might be involved in a bidirectional transcription regulation of the YP2 and the tubulin gene.

The YP2 gene copy number was determined with a ³²P-nick-translated 526-bp *Eco*RI/*Hinc*II fragment (including codons 16–190) which was hybridized to restriction endonuclease-digested *S. cerevisiae* DNA on Southern blots. As shown in Fig. 4, only one hybridizing fragment was detected with either *Bam*HI-, *Hind*III-, *Eco*RI- or *Pst*I-digested DNA. The lengths of these fragments were those predicted from the known location of restriction sites within or adjacent to the YP2 gene (see Fig. 1). As the *Pst*I cutting site lies in the coding segment used as hybridization probe, two hybridizing *Pst*I fragments were expected. Only the known 5.1-kb fragment cloned in the plasmid pYA208 (see Fig. 1) displayed a significant hybridization signal, but longer exposure of the autoradiogram shown in Fig. 4 gave a second hybridizing fragment of about 11 kb (from which only 145 bp were contained in the hybridization probe). As the same Southern blot results were obtained in stringent and less stringent filter washing conditions (65 °C and either 0.1 × SSC or 1 × SSC), we conclude that there is only one copy of the YP2 gene per haploid genome.

Since any two proteins with more than, say, 30% sequence homology over more than 100 residues have essentially equal secondary and tertiary structure³³, we propose that the YP2 protein of yeast and the *c-ras* and *v-ras* p21 proteins of vertebrates have similar function, but perhaps altered specificity. The functions of p21 proteins are not known, but it has been established that they bind GTP^{22–24}. Since YP2 is strongly conserved relative to the *ras* gene products in the region predicted to contribute to nucleotide binding^{28,29}, YP2 is also likely to bind GTP. However, binding motifs of YP2 are far from perfectly

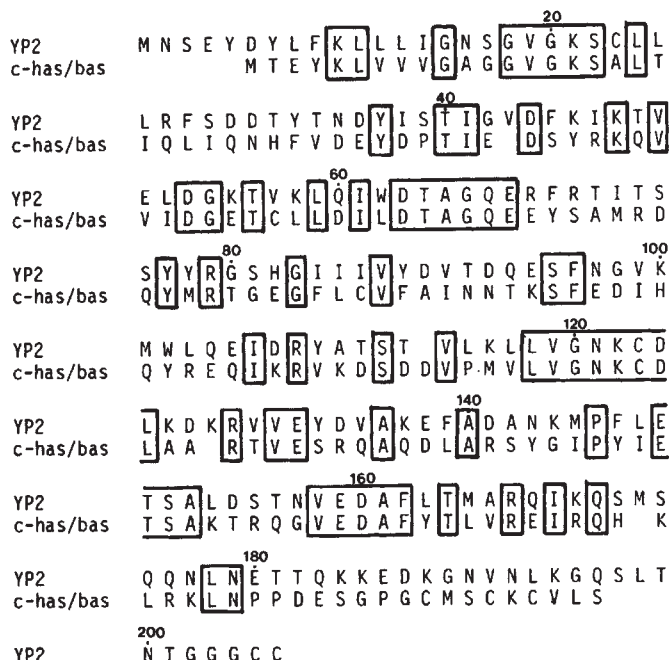


Fig. 5 Alignment of the amino acid sequences of the YP2 and the p21 protein as deduced from the nucleotide sequences of the respective genes. The human *c-has/bas* sequence is from refs 4 and 5. Identical residues are boxed. The residues of the YP2 protein are numbered beginning with the initiating methionine (see Fig. 3). Sequences are given in the one letter code.

matched to the consensus sequence. For example, there is sequence agreement with the predicted nucleotide-binding site of bovine ATPase (β -subunit), but it is essentially limited to Lys-Leu-X-Leu-X-Gly-X-X-Gly-Val-Gly-Lys-Ser where Val/Leu and Ser/Thr are counted as identities. The best working hypothesis for mutational tests of nucleotide binding remains, however, the region around Gly 12 (p21 *ras*)/Ser 17 (YP2). Gln 61 (p21 *ras*)/Gln 67 (YP2) may also be near the nucleotide-binding site in the tertiary structure, consistent with the observed point mutation of the activated Hs242 *c-has/bas*²¹.

Focusing on the two transformational hotspots of p21, it is interesting that YP2 protein shares a glutamine 61 with human proto-oncogene derived p21 but shares a serine 12 with the transforming *v-ras*^K p21, suggesting that normal cellular function is possible in spite of the Gly → Ser replacement, contrary to the prediction of Wierenga and Hol²⁹. Alternatively, the functional specificity of the YP2 protein in fast-growing yeast cells may be like that of the viral oncogene-derived protein in transformed vertebrate cells.

As mentioned above, the inverted repeat structure found about 80 bp in front of the 5' start sites of the YP2 and the tubulin mRNAs might be involved in a bidirectionally acting common regulation of their two genes. Taking this as a working hypothesis (which is testable by mutating or deleting this sequence segment), one function of the YP2 could be viewed as a GTP-binder interacting with tubulin in the tubulin assembly process which is GTP-dependent (for review see ref. 30). Alternatively, it could function as a phosphokinase or as part of a ATPase or GTPase, as suggested by viral p21s on the basis of sequence homology with *Escherichia coli* ATPase β -subunit²⁸. In this context it is worth mentioning that viral p21s autophosphorylate the threonine in position 59²².

Whatever the biological function of this protein, its conservation from yeast to man points to it having an elementary physiological role in all eukaryotic cells. Since there is only one copy of the YP2 gene in the yeast genome which lends itself to relatively easy genetic manipulation, there is hope that the biochemical function of the protein can be elucidated in this unicellular organism.

We thank Renate Seidel, Angela Fiebiger and Petra Teutsch for technical assistance and Franz-Josef Klinz and Christopher Langford for critically reading the manuscript. This work was supported by grants to D.G. from the Deutsche Forschungsgemeinschaft, the Kempkes-Stiftung of the University of Marburg and the Fonds der Chemischen Industrie.

Received 15 August; accepted 16 September 1983.

- Shilo, B. Z. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6789-6792 (1981).
- Hoffman-Falk, H., Einat, P., Shilo, B. Z. & Hoffman, F. M. *Cell* **32**, 589-598 (1983).
- Simon, M. A., Kornberg, T. B. & Bishop, J. M. *Nature* **302**, 837-839 (1983).
- Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. *Nature* **302**, 33-37 (1983).
- Reddy, E. P. *Science* **220**, 1061-1063 (1983).
- Dhar, R. *et al. Science* **217**, 934-937 (1982).
- Tsuchida, N., Ryder, T. & Ohtsubo, E. *Science* **217**, 937-939 (1982).
- Gallwitz, D. & Seidel, R. *Nucleic Acids Res.* **8**, 1043-1059 (1980).
- Gallwitz, D. & Sures, I. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2546-2550 (1980).
- Neff, N. F., Thomas, J. H., Grisafi, P. & Botstein, D. *Cell* **33**, 211-219 (1983).
- Mortimer, R. K. & Schild, D. in *The Molecular Biology of the Yeast *Saccharomyces cerevisiae*. Metabolism and Gene Expression* (eds J. N. Strathern, E. W. Jones & J. R. Broach) 639-650 (Cold Spring Harbor Laboratory, New York, 1982).
- Gallwitz, D., Perrin, F. & Seidel, R. *Nucleic Acids Res.* **9**, 6339-6350 (1981).
- Maxam, A. & Gilbert, W. *Meth. Enzym.* **65**, Pt 1, 499-560 (1980).
- Hartwell, L. H., McLaughlin, C. S. & Warner, J. R. *Molec. gen. Genet.* **109**, 42-56 (1970).
- Rosbash, M., Harris, P. K. W., Woolford, J. & Teem, J. L. *Cell* **24**, 679-686 (1981).
- Langford, C. & Gallwitz, D. *Cell* **33**, 519-527 (1983).
- Smith, T. F. & Waterman, M. S. *J. molec. Biol.* **147**, 195-197 (1981).
- McLachlan, A. D. *J. molec. Biol.* **61**, 409-424 (1971).
- Tabin, C. J. *et al. Nature* **300**, 143-149 (1982).
- Reddy, E. P., Reynolds, R. K., Santos, E. & Barbacid, M. *Nature* **300**, 149-152 (1982).
- Yusa, Y. *et al. Nature* **303**, 775-779 (1983).
- Scolnick, E. M., Papageorge, A. G. & Shih, T. Y. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5355-5359 (1979).
- Shih, T. Y., Papageorge, A. G., Stokes, P. E., Weeks, M. O. & Scolnick, E. M. *Nature* **287**, 686-691 (1980).
- Papageorge, A., Lowy, D. & Scolnick, E. M. *J. Virol.* **44**, 509-519 (1982).
- Berk, A. J. & Sharp, P. A. *Cell* **12**, 721-732 (1977).
- Weaver, R. & Weissmann, C. *Nucleic Acids Res.* **7**, 1175-1193 (1979).
- Tinoco, I. *et al. Nature new Biol.* **246**, 40-41 (1973).
- Gay, N. J. & Walker, J. E. *Nature* **301**, 262-263 (1983).
- Wierenga, P. K. & Hol, W. G. J. *Nature* **302**, 842-844 (1983).
- Timasheff, S. N. & Grisham, L. M. *A Rev. Biochem.* **49**, 565-591 (1980).
- Maniatis, T., Jeffrey, A. & van de Sande, H. *Biochemistry* **14**, 3787-3784 (1975).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
- Kabsch, W. & Sander, C. *Biopolymers* (in the press).

ras*-Related gene sequences identified and isolated from *Saccharomyces cerevisiae

D. DeFeo-Jones*, E. M. Scolnick*, R. Kollert† & R. Dhar†

* Department of Virus and Cell Biology, Merck Sharp & Dohme Research Laboratories, West Point, Pennsylvania 19486, USA
† Laboratory of Molecular Virology, National Cancer Institute, Bethesda, Maryland 20205, USA

The oncogenes of Harvey and Kirsten murine sarcoma viruses (*v-ras^H* and *v-ras^K*) and their cellular homologues (*c-ras^H* and *c-ras^K*) constitute two members of the *ras* gene family. Each functional member of the *ras* gene family encodes a 21,000 molecular weight protein (p21^{ras})¹. *ras* genes have been detected in a wide variety of vertebrate species², including *Xenopus laevis* (R. E. Steele, personal communication), and in *Drosophila melanogaster*³. We report here the detection of *ras*-related genes in the yeast *Saccharomyces cerevisiae*, and the isolation of two *ras*-related molecular clones, *c-ras^{sc-1}* and *c-ras^{sc-2}*, from the DNA of *Saccharomyces*. Both *c-ras^{sc-1}* and *c-ras^{sc-2}* hybridize specifically to probes prepared from mammalian *ras* DNA. Sequencing of *c-ras^{sc-1}* reveals extensive amino acid homology between the protein encoded by *c-ras^{sc-1}* and the p21 encoded by *c-ras^H*. Our studies suggest that these clones can be used to elucidate the normal cellular functions of *ras*-related genes in this relatively simple eukaryotic organism.

The transforming genes of Harvey and Kirsten murine sarcoma viruses have been named *ras* oncogenes because these two oncogenic retroviruses were originally isolated by passage of murine leukaemia viruses in rats¹. The *ras* oncogene of Harvey

Fig. 1 Southern blot analysis of restriction endonuclease-digested high molecular weight DNA prepared from the yeast *S. cerevisiae*¹². Haploid genomic DNA (20 µg) was digested with the enzymes *Eco*RI (lane 1) and *Xba*I (lane 2), electrophoresed in agarose gels, transferred to nitrocellulose filters and hybridized to ³²P-labelled nick-translated probes prepared from either *v-ras^H*-specific cloned DNA (plasmid clone 14)¹³ (a) or the yeast-derived *ras^H*-reaction clone *c-ras^{sc-1}* (b) (this clone was isolated as described in Fig. 2 legend). The autoradiogram in a represents an 18-h exposure. Approximate fragment sizes are indicated in kbp.

Methods: Hybridizations were performed at 42 °C for 24-36 h in a solution containing 40% formamide; 6×SSC (1×SSC = 0.015 M sodium citrate, 0.15 M sodium chloride), 4×Denhardt's containing 0.8% each bovine serum albumin, Ficoll and polyvinylpyrrolidone; 0.1% SDS; and 1 mg ml⁻¹ low molecular weight salmon testis DNA. The filters were washed sequentially, three times each for 20 min, in 2×SSC, 1×SSC and 0.5×SSC with 0.1% SDS at 22 °C and 42 °C, respectively.

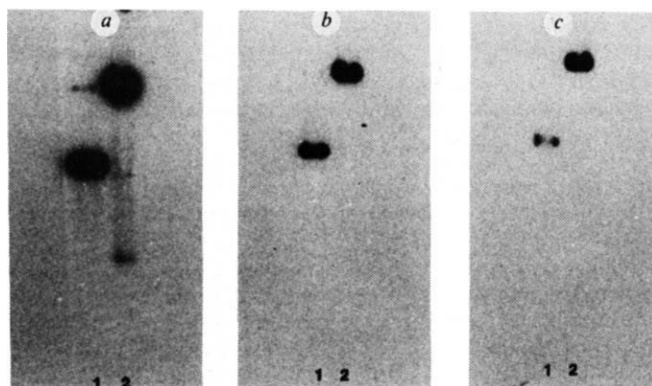
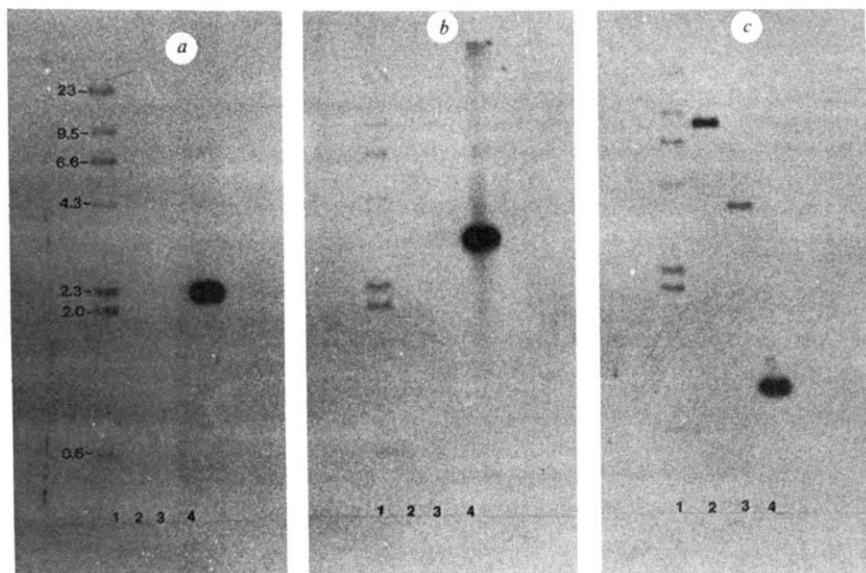


Fig. 2 Effect of hybridization conditions on homology between *v-ras^H* and yeast clones *c-ras^{sc-2}* and *c-ras^{sc-1}*. The two *ras^H*-reactive yeast clones, *c-ras^{sc-2}* (lane 1) and *c-ras^{sc-1}* (lane 2), were isolated from a partial *Eco*RI library of yeast DNA cloned into the bacteriophage vector λgtw10AB (gift of Dr L. Schultz). a, 55 °C hybridization, 60 °C washes. b, 60 °C hybridization, 60 °C washes. c, 60 °C hybridization, 65 °C washes. **Methods:** Phage DNAs were prepared according to published procedures¹⁴, digested with the restriction endonuclease *Eco*RI to release the inserts, electrophoresed on agarose gels, transferred to nitrocellulose filters and hybridized to ³²P-labelled *v-ras^H*-specific cloned DNA. The hybridization solution (4×SSC, 1×Denhardt's, 0.1% SDS and 1 mg ml⁻¹ low molecular weight DNA) and wash solutions (1×SSC, 0.1% SDS and 0.1×SSC and 0.1% SDS, respectively) used in each case were the same; however, the temperature of hybridization and/or washes was elevated in order to increase the stringency conditions.

murine sarcoma virus has been named *v-ras^H* and that of Kirsten murine sarcoma virus has been named *v-ras^K* (ref. 4). Recently, mutated forms of *c-ras^H* and *c-ras^K* have been identified in certain human cancers and in cell lines derived from these cancers⁵. An additional member of the *ras* gene family, called *n-ras*, has also been found as an activated oncogene in a human neuroblastoma⁶.

A molecularly cloned 1.0-kilobase pair (kbp) *v-ras^H*-specific DNA probe was used in relaxed hybridization conditions to detect *ras*-related sequences by Southern blot analyses of *Eco*RI- and *Xba*I-digested yeast genomic DNA (Fig. 1a, lanes 1, 2). Two major bands of approximately 8.0 and 3.8 kbp were detected in the *Eco*RI-digested DNA. A minor band of approximately 1.8 kbp was also detected. However, the hybridization of *v-ras^H* to this minor band has not been consistently reproducible.

Fig. 3 Southern blot analysis showing specificity of hybridization between ^{32}P -labelled *v-ras^H*-specific DNA and the bacteriophage yeast clones *c-ras^{sc}-1* and *c-ras^{sc}-2*. Lane 1 in each panel represents ^{32}P -labelled λ DNA markers in kbp. In panels *a*, *b* and *c*, three different ^{32}P -labelled probes are used. *a*, A ^{32}P -labelled probe prepared from DNA representing a segment of the glycoprotein D (*gD*) gene from herpes simplex I virus was used in hybridizations against 1 μg each of *c-ras^{sc}-1* (lane 2), *c-ras^{sc}-2* (lane 3) and the *gD* gene segment from which the probe was made (lane 4). *b*, ^{32}P -labelled probe was prepared from *v-src* DNA and hybridized to 1 μg each of *c-ras^{sc}-1* (lane 2), *c-ras^{sc}-2* (lane 3) and the *v-ras* gene segment from which the probe was made. *c*, A *v-ras^H*-specific probe was used in hybridizations against 1 μg each of *c-ras^{sc}-1* (lane 2) and *c-ras^{sc}-2* (lane 3). As a means of estimating nucleic acid divergence between *ras^H* and the two yeast clones, 1 μg of *v-ras^K*-specific DNA was used as the positive control (lane 4). All hybridizations and washes were performed at 60 °C in the same conditions as those described in Fig. 2 legend. The ^{32}P -labelled probes used were prepared from DNA fragments which were isolated from plasmid clones by gel electrophoresis before nick translation.



The *Xba*I-digested DNA showed three hybridizing bands of approximately 4.3, 2.0 and 1.4 kbp. Having identified these *ras*-homologous sequences, we used a *v-ras^H* probe to screen a bacteriophage library generated from a partial *Eco*RI digest of yeast DNA (see Fig. 2 legend) and were able to isolate two *Eco*RI clones of 8.0 and 3.8 kbp (*c-ras^{sc}-1* and *c-ras^{sc}-2*, respectively). The sizes of these clones correspond to the two major *Eco*RI fragments detected in the genomic DNA (Fig. 1a, lane 1).

Figure 2 shows that *v-ras^H* continued to hybridize to both *c-ras^{sc}-1* and *c-ras^{sc}-2* in increasingly stringent temperature conditions. The same result was achieved using a *v-ras^K*-specific probe, HiHi 380⁷ (data not shown). To investigate the specificity of this hybridization to *v-ras^H* further, we performed hybridizations to *c-ras^{sc}-1* and *c-ras^{sc}-2* using probes made from other molecularly cloned DNAs: (1) a segment of the glycoprotein D (*gD*) gene of herpes simplex I virus (gift from Pat Spear) (Fig. 3a), (2) the *src* gene of Rous sarcoma virus (Fig. 3b) and (3) *v-ras^H*-specific DNA (Fig. 3c). The results clearly show that the hybridization is specific for *ras* sequences. Because the herpes *gD* gene is G+C rich⁸, the lack of hybridization with this probe eliminates the possibility that the *ras* homology seen is the result of spurious hybridization due to long stretches of these nucleotides in the yeast clones. A comparison of the intensity of hybridization of *v-ras^H* (Fig. 3c) to *c-ras^{sc}-1* and -2 (lanes 2, 3) and the positive control *v-ras^K* (lane 4) suggests that *c-ras^{sc}-1* and -2 share less than 75% homology to *v-ras^H* (refs 9, 10). In addition, on the basis of this comparison, it appears that *c-ras^{sc}-2* may contain less *ras* information or may be more divergent from *v-ras^H* than *c-ras^{sc}-1*. To determine the copy number of *ras*-related genes in the yeast genome, *c-ras^{sc}-1* was subcloned into the *Eco*RI site of PBR322 (see Fig. 4 legend) and used to probe *Eco*RI- or *Xba*I-digested yeast genomic DNA (Fig. 1b, lanes 1, 2). Two bands were detected in the *Eco*RI digest; their sizes (8.0 and 3.8 kbp) correlate well with the sizes of *c-ras^{sc}-1* and *c-ras^{sc}-2* (Fig. 1, lane 1).

In the *Xba*I-digested DNA, we detected three major bands of 4.3, 2.0 and 1.4 kbp (Fig. 1b, lane 2), respectively. These sizes match well with those of the three major *Xba*I bands which hybridize to *v-ras^H* (Fig. 1a, lane 2). Restriction mapping of *c-ras^{sc}-1* revealed the presence of three very closely spaced *Xba*I sites towards the middle of the *ras* homologous sequences (Fig. 4). These sites partition this clone into two major *ras*-reactive *Xba*I fragments so that the presence of three major *Xba*I fragments in the yeast genomic DNA strongly suggests the presence of two *ras*-related genes in yeast. The less intensely

hybridizing band of ~3.3 kbp in the *Xba*I-digested DNA is not reproducible and probably represents a partial digest.

A restriction map of *c-ras^{sc}-1* was constructed and the region of *ras* homology determined (Fig. 4a). A single *Hind*III subfragment of ~1.8 kbp was found to contain all the *ras*-homologous sequences (Fig. 4a, barred areas). This fragment was subcloned into the *Hind*III site of pBR322 and Maxam-Gilbert sequencing was performed. A single open reading frame was found in the *ras*-related region of *c-ras^{sc}-1*. The amino acid sequence deduced from the nucleotide sequence of *c-ras^{sc}-1* and the comparative amino acid sequence of *c-ras^H* are shown in Fig. 4b. The *ras* gene from *S. cerevisiae* begins with seven additional amino acids at the N-terminus of the protein coded for by this gene. Thus, in Fig. 4b the amino acids coded for by *c-ras^{sc}-1* and *c-ras^H* have been aligned beginning at position 1 of *c-ras^H* and position 8 of *c-ras^{sc}-1*.

Extensive stretches of identical amino acids are coded for by the DNA sequences of *c-ras^H* and *c-ras^{sc}-1* in each of the first three coding exons of *c-ras^H*. The amino acid homology is closest up to position 125 of *c-ras^H* and is more divergent between positions 125 and 158. Certain features of the amino acid sequences are of interest. (1) Position 12 of the *c-ras^H* and position 19 of the yeast *ras* gene have a glycine residue. (2) Position 59 of *c-ras^H* has an alanine and the homologous position 66 of *c-ras^{sc}-1* has an alanine. (3) Position 61 of *c-ras^H* has a glutamine and position 68 of *c-ras^{sc}-1* has a glutamine. The amino acids at these positions in *ras* genes have been noted to be altered in various oncogenic *ras* genes isolated from rodent and human cells^{1,5,11} and it will be important to test the effect of comparable amino acid changes on the activity of *c-ras^{sc}-1* expressed in both yeast cells and in mammalian cells. The DNA sequencing of *c-ras^{sc}-1* (and *c-ras^{sc}-2*) is not sufficiently secure beyond the data presented in Fig. 4b. A detailed complete sequence of both genes will be presented elsewhere. Northern blot hybridizations using probe made with *c-ras^{sc}-1* against either total or poly(A)-selected yeast RNA reveals a single transcript, indicating that this gene is being actively transcribed. The current data are clearly sufficient to indicate that *S. cerevisiae* contains at least one *ras*-like gene. Presumably, the amino acid homologies between genes from such evolutionarily divergent organisms reflect conservation of regions of the *ras* proteins required for common functions of the proteins. Thus, a detailed structural and genetic analysis of how these *ras* gene sequences function in yeast may open the way for a unique understanding of the functional role(s) of oncogenes in both normal and malignant higher eukaryotic cells.

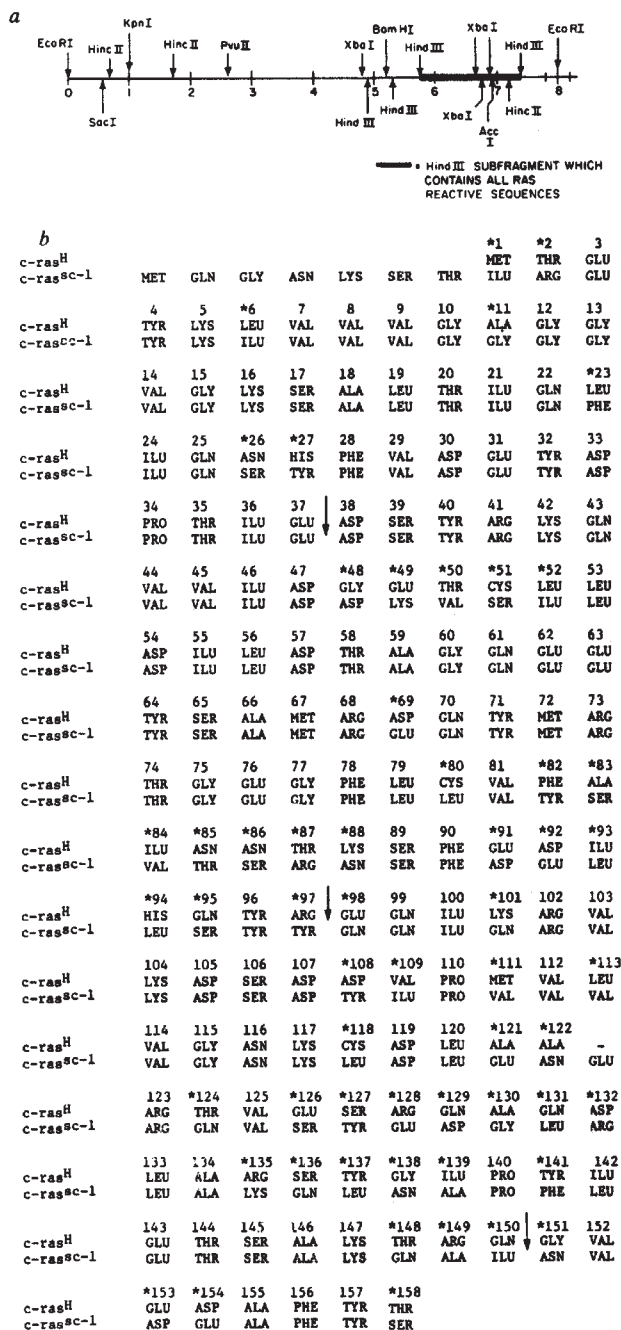


Fig. 4 *a*, Restriction endonuclease map of yeast clone *c-ras*^{sc-1}. The cloned yeast insert DNA was electrophoretically isolated from its bacteriophage vector λ gtwes λ B by digestion with the restriction enzyme *Eco*RI. This fragment was then subcloned into the *Eco*RI site of the plasmid pBR322. The plasmid clone was grown and DNA prepared according to published procedures¹⁴. Total *c-ras*^{sc-1} plasmid DNA and/or the purified *c-ras*^{sc-1} insert were used in restriction endonucleases analyses. 1–2 μ g of DNA were digested with various enzymes either singly or in combination and then gel electrophoresed. The sizes of the resultant fragments were determined by comparison with *Hind*III-digested λ marker. Restriction sites are given in kbp and are accurate to ± 0.1 kbp. The *v-ras*-reactive region of *c-ras*^{sc-1} is represented as a solid block contained within a single *Hind*III subfragment. This information was obtained by Southern blot hybridization of the restriction fragments of *c-ras*^{sc-1} with ³²P-labelled *c-ras*^H probe. *b*, Amino acid sequence homology of *c-ras*^{sc-1} and *c-ras*^H. Maxam–Gilbert sequence analyses were performed on the *ras*-homologous *Hind*III subfragment (*a*) of *c-ras*^{sc-1} as previously described⁹. The upper line represents the amino acid sequence of *c-ras*^H and the lower line that of *c-ras*^{sc-1}. *c-ras*^{sc-1} contains seven additional N-terminal amino acids. The arrows indicate positions where introns occur in the nucleotide sequence of *c-ras*^H. The asterisks denote those amino acids which are different between *c-ras*^H and *c-ras*^{sc-1}.

We thank R. W. Ellis, R. E. Jones and G. Temeles for helpful discussions during the preparation of this manuscript.

Received 5 September; accepted 28 September 1983.

1. Ellis, R. W., Lowy, D. R. & Scolnick, E. M. *Adv. viral Oncol.* **1**, 107-126 (1982).
2. Weiss, R. A., Teich, N. M., Varmus, H. E. & Coffin, J. M. (eds) *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, New York, 1982).
3. Shilo, B. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6789-6792 (1981).
4. Coffin, J. M. *et al. J. Virol.* **40**, 953-957 (1981).
5. Cooper, G. M. *Science* **218**, 801-804 (1982).
6. Shimizu, K., Goldfarb, M., Perucho, M. & Wigler, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 387-391 (1983).
7. Ellis, R. W. *et al. Nature* **292**, 506-511 (1981).
8. Watson, R. J., Weiss, J. H., Salstrom, J. S. & Enquist, L. W. *Science* **218**, 381-384 (1982).
9. Dhar, R. *et al. Science* **217**, 934-937 (1982).
10. Tsuchida, N., Ryder, T., & Ohtsubo, E. *Science* **217**, 937-938 (1982).
11. Yuasa, J. *et al. Nature* **303**, 775-779 (1983).
12. Wickner, R. B. & Leibowitz, M. J. *Genetics* **82**, 429-??? (1976).
13. Ellis, R. W. *et al. J. Virol.* **36**, 408-420 (1980).
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. Jr *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

Selective unfolding of erythroid chromatin in the region of the active β -globin gene

**Takeshi Kimura, Frederick C. Mills, James Allan
& Hannah Gould**

Department of Biophysics, King's College, London,
26-29 Drury Lane, London WC2B 5RL, UK

Globin gene expression, which occurs exclusively in the erythroid cell lineage, is controlled at the level of transcription¹. It is thus of some considerable interest to compare the chromatin structure of this gene with that of inactive genes in erythroid cell nuclei and to compare the chromatin structure of the globin gene in its active and inactive states in nuclei of different cell types. Other workers have observed that globin genes in erythroid cell nuclei exhibit the enhanced overall sensitivity to nucleases² and the hypersensitive site in the 5'-flanking sequence³ typical of many active genes⁴. The nature of the structural changes giving rise to nuclease sensitivity are however obscure. We have investigated the local higher order structure of chromatin in the region of unique genes in chicken by sucrose gradient sedimentation of chromatin restriction fragments. We find that ovalbumin and $\alpha 2$ -collagen gene fragments in erythrocyte chromatin and an adult β -globin gene fragment in spleen chromatin sediment with bulk chromatin fragments of the same DNA size, whereas the β -globin gene fragment in erythrocyte chromatin sediments more slowly than bulk fragments of equivalent size. The simplest interpretation of the results is that the solenoid structure in the region of the globin gene is selectively and permanently unfolded on gene activation.

We have previously shown that linker histone (H1 and H5) induced chromatin condensation, as measured by electron microscopy, is directly proportional to the sedimentation rate⁵. Here we have used the sedimentation rate of chromatin as the primary measure of its higher order folding. Our experimental approach involves digestion of isolated nuclei with the restriction endonuclease *Eco*RI, followed by sucrose gradient fractionation of chromatin fragments and analysis of DNA in individual fractions by electrophoresis through an agarose gel. Ethidium bromide staining of the gel reveals the size distribution of bulk DNA in fractions of the gradient, from which distributions of fragments of any selected size, in particular those identical in size to specific gene fragments of interest, can be derived. After transfer of the DNA to nitrocellulose, comparable distributions of specific gene fragments can be determined by hybridization with labelled probes and autoradiography⁶.

Fig. 1 Distribution of *Eco*RI chromatin fragments from chicken erythrocyte chromatin in sucrose gradient fractions. *a*, Absorbance profile of *Eco*RI fragments in the sucrose gradient. *b*, *Eco*RI restriction maps of ovalbumin¹⁴, globin¹⁵ and collagen¹⁶ gene regions. Solid lines below each map represent the main hybridising regions for each probe, for ovalbumin pOV230¹⁷, for globin pCABG1subEH, subcloned from λ vesCABG1¹⁸, and for collagen p α 2CG71sub68¹⁹. Dotted lines below represent cross-hybridising regions. D, E and F represent *Eco*RI fragments seen in *d*, *f* and *g*. Panels *c* and *e* are fluorescence photographs of DNA fractionated by agarose gel electrophoresis and stained with ethidium bromide. *d*, Autoradiograph of Southern blot of gel *c* probed for ovalbumin, showing fragments A, B, C (see text) and D. *f*, Autoradiograph of gel *e* probed for globin, showing fragment E. Smaller bands arose from the hybridisation of repetitive elements due to the use of *Escherichia coli* rather than salmon sperm DNA in the hybridization. *g*, Autoradiograph of the same blot as *f*, reprobed for collagen, showing collagen fragment F and globin fragment E. Fluorescence photographs and autoradiographs were enlarged photographically to the same size to align the corresponding positions. Numbers below panels represent fraction numbers and M represent marker lanes. Markers were λ phage *Eco*RI fragments (³²P-labelled and unlabelled) and λ Charon4A phage *Bam*HI fragments (unlabelled) whose sizes are given beside panels. M bands in *f* were clearly visible on longer exposures.

Methods: Blood from a White Leghorn chicken was collected through two layers of surgical gauze into a solution of cold phosphate-buffered saline (PBS) with 5 mM sodium butyrate, 1 mM EGTA, 0.2 mM phenylmethylsulphonyl fluoride (PMSF) and 25 units ml⁻¹ heparin. All subsequent steps were performed on ice or at 4 °C. Blood was sedimented at 2,000g for 5 min and washed 3 times in PBS with 5 mM sodium butyrate, 1 mM EGTA and 0.2 mM PMSF. After removal of the supernatant and white cell layer, red blood cells were resuspended in a minimum volume of 0.25 M sucrose, 6 mM MgCl₂, 50 mM Tris-HCl (pH 7.5), 5 mM sodium butyrate, 1 mM EGTA and 0.2 mM PMSF (buffer A), then added dropwise to a large volume (~ ×10) of buffer A with 1% Triton X-100 and stirred rapidly for 17 min. Nuclei were sedimented as described above and washed 3–5 times with buffer A, then 3–5 times with 33 mM NaCl, 5 mM MgCl₂, 50 mM Tris-HCl (pH 7.5), 5 mM sodium butyrate, 1 mM EGTA and 0.2 mM PMSF. 9 volumes of nuclei were digested by 1 vol. of *Eco*RI stock (Boehringer), with the respective final concentrations of approximately 10 mg ml⁻¹ and 10,000 unit ml⁻¹, for 2–3 h at 37 °C. The reaction was stopped and nuclei were lysed with 4 vol. of an ice-cold solution of 87.5 mM NaCl, 1.5 mM EDTA, 5 mM sodium butyrate and 0.2 mM PMSF, and left overnight at 4 °C (sodium butyrate made lysis difficult). Nuclear debris was pelleted at 2,000g for 5 min and the supernatant was used for sedimentation experiments. Integrity of histones was monitored by polyacrylamide gel electrophoresis¹⁹. Preparations from chicken erythrocytes and spleen (30 min digest) were satisfactory; those from reticulocytes or fibroblasts were degraded. Samples containing 10–20 absorbance units (260 nm) of the *Eco*RI chromatin fragments were sedimented through linear sucrose gradients (20–50% sucrose, 80 mM NaCl, 10 mM Tris-HCl (pH 7.5), 5 mM sodium butyrate, 0.1 mM EDTA, 0.2 mM PMSF), at 82,000g for 2.5 h at 10 °C. Fractions were collected and DNA was isolated and 5 μ g were electrophoresed through 0.6% (*c*) or 0.8% (*e*) horizontal agarose gels. Gels were stained with ethidium bromide and fluorescence under ultraviolet illumination photographed. DNA from these gels was transferred onto nitrocellulose filters essentially according to Southern⁶. Filters were pre-hybridized with 3 × SSC, 0.2% (w/v) each of bovine serum albumin, Ficoll and polyvinylpyrrolidone, 0.2% SDS, 5 mM EDTA, 0.5 mg ml⁻¹ single-stranded sonicated salmon sperm (Figs 1*d*, *g* and 4*c*) or *E. coli* (Fig. 1*f*) DNA for over 2 h at 65 °C and then hybridized to a ³²P-labelled specific DNA probe²⁰ (1–5 × 10⁸ d.p.m. μ g⁻¹) in pre-hybridization buffer with 9% (w/v) dextran sulphate for over 12 h at 65 °C. Filters were washed extensively in 2 × SSC and 0.1% SDS followed by washing in 0.1 × SSC and 0.1% SDS at 65 °C, and then dried and autoradiographed with intensifying screens (*d*, *f*). The filter in *f* was washed in distilled water at 70 °C and then rehybridized to the labelled collagen probe as above, but with 5 μ g ml⁻¹ single-stranded sonicated chicken DNA added to reduce the signal from the repetitive element (*g*).

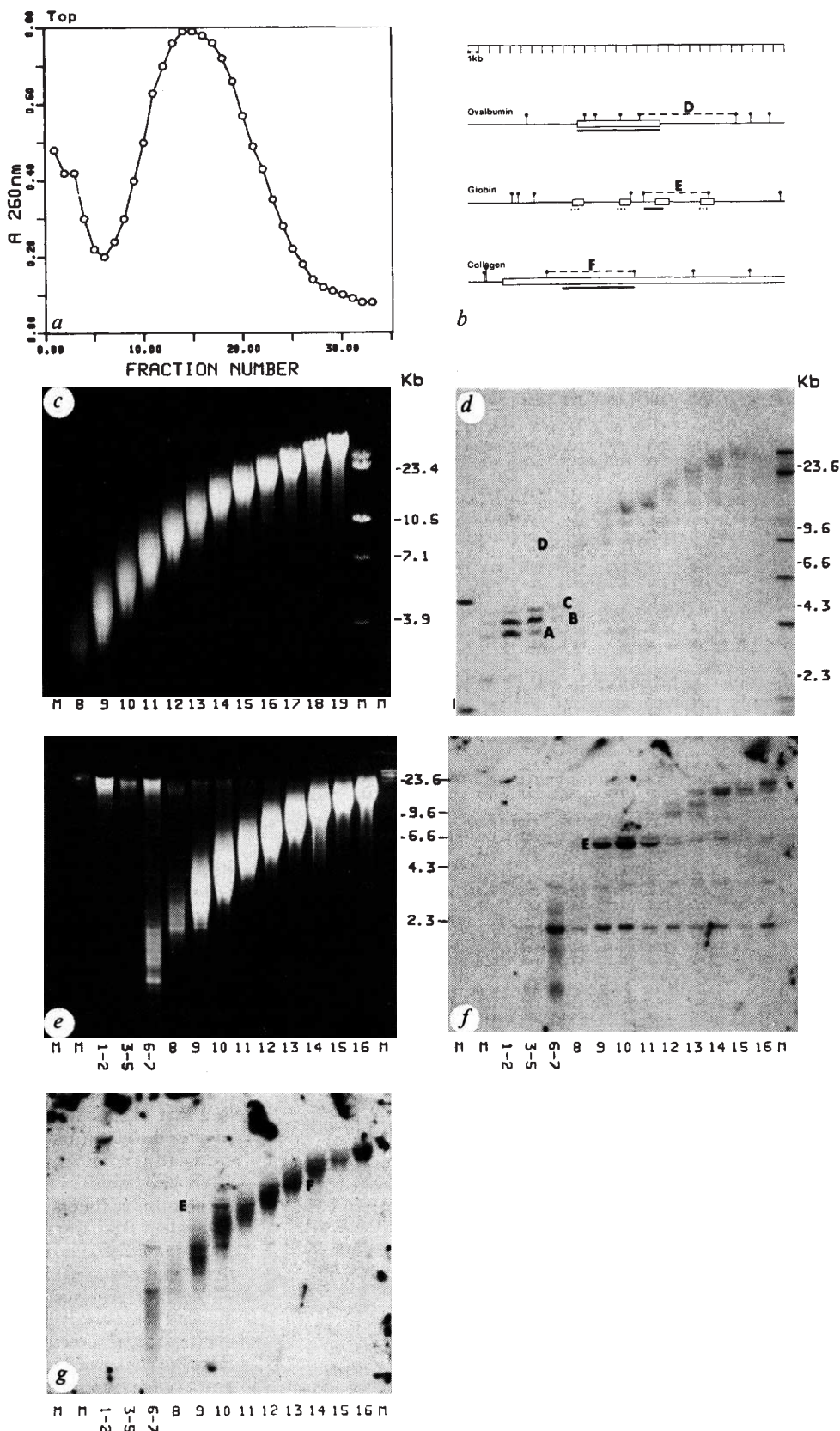


Fig. 2 Typical analysis of distributions of total DNA and specific gene sequences. Photographic negatives of ethidium bromide fluorescence and autoradiographs of Fig. 1c-g were scanned with a Joyce Loebel KkIIIc microdensitometer. Marker positions (and sometimes an image of a ruler on the negative) were used to obtain the size conversion factors between the corresponding traces. *a*, Densitometer traces from gel lanes for fractions 8-11 of the photographic negative of Fig. 1c (upper trace) and the autoradiograph of Fig. 1d (lower trace). *b*, Densitometer traces for fractions 9-12 of Fig. 1e and Fig. 1f. Dots on the upper traces mark fluorescence intensities at the peak positions of ovalbumin fragments A, B, C and globin fragment E. Relative intensities of the specific gene fragments seen in the autoradiographic profiles and those at equivalent (dotted) positions on the bulk DNA profiles were used to construct the histograms in Fig. 3.

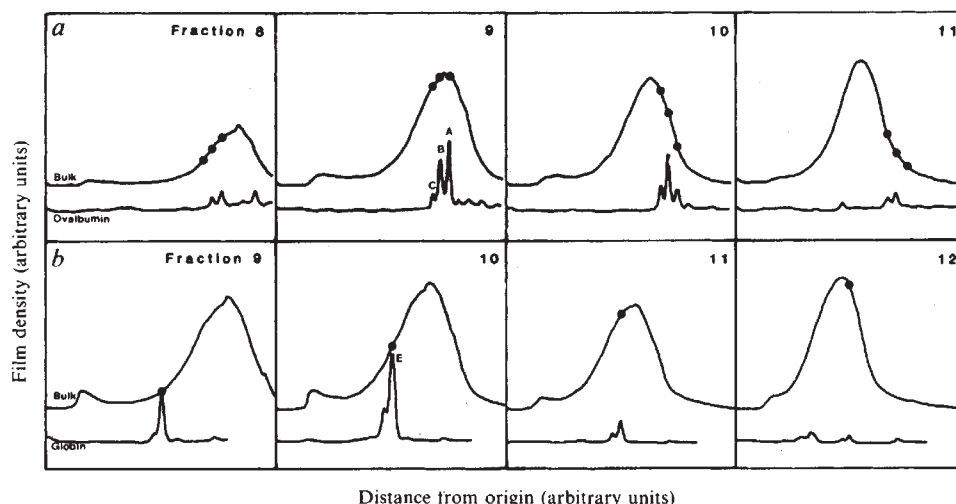
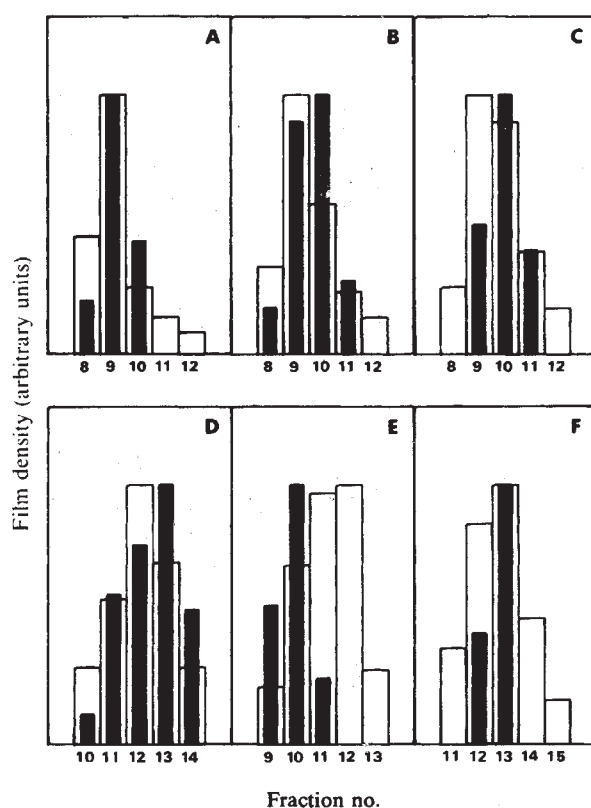


Fig. 3 Histograms representing sedimentation of specific gene fragments and bulk DNA of the same size. Data obtained as described in Fig. 2 were normalized so that the maxima for specific and bulk DNA distributions in each panel had the same height. Filled bars represent specific gene fragments and open bars bulk DNA of equivalent size. Letters correspond to respective specific gene fragments seen previously in Figs 1 and 2 and so histograms A-D are for ovalbumin gene fragments, E for the globin gene fragment and F for the collagen gene fragment. Essentially the same results were obtained when the photographic response of the films were taken into account. DNA from fractions 8-13 was analysed by electrophoresis on a low percentage agarose gel (not shown) and, using a computer program package developed by T.K. and F.M., were found to have the number average sizes of 2.8, 3.7, 4.9, 6.2, 7.9 and 10.0 kb. The maximum of the 6.2 kb globin fragment distribution corresponds to bulk DNA with an average size of 4.9 kb, while the maxima of the 9.3 kb ovalbumin and 8.3 kb collagen fragments correspond to bulk DNA with an average size of 10.0 kb.



Mature chicken erythrocytes afford a convenient source of chromatin in which to examine localised differences in higher order structure. Globin genes have been expressed in this lineage, but not in that of spleen cells, which are used for comparison. Ovalbumin and collagen are taken to represent genes that are not expressed in erythroid cells. Although RNA synthesis⁷, and indeed globin mRNA synthesis⁸, is largely blocked in the mature erythrocyte, nuclease sensitivity of the globin gene⁹ and hypersensitivity of the 5'-flanking sequence³ indicate that at least these characteristics of active chromatin in the less mature erythroblasts^{2,3} are maintained.

Figure 1 shows the sedimentation of bulk and specific *EcoRI* fragments of erythrocyte chromatin. As can be seen from Fig. 1d, f, g, chromatin fractions contain specific fragments of globin, ovalbumin and collagen genes, representing partial *EcoRI* digests. DNA sizes measured from gel markers do not correspond exactly with sizes deduced from the maps in Fig. 1b owing to the overloading required for the detection of fragments. By referencing approximate sizes of the fragments to the restriction maps, it can be deduced that band D in Fig. 1d is the 9.3 kilobase (kb) ovalbumin fragment and band E in Fig. 1f is the 6.2 kb globin fragment, while bands A, B and C in Fig. 1d must come from the 5' half of the ovalbumin gene. The collagen gene probe contained a repetitive element which is responsible for the broad background underlying the specific 8.3 kb collagen fragment F in fig. 1g. Hybridization of the labelled probe to repetitive elements was minimised by inclusion of unlabelled chicken DNA in the hybridization mixture. A trace of globin band E remains from the first hybridization. The presence of repetitive elements, which have the same distribution as bulk DNA, and the globin band E allows a direct comparison with the collagen band F.

Electrophoretic profiles of bulk DNA and specific gene fragments in individual sucrose gradient fractions were analysed by

densitometry as shown in Fig. 2. The relative amount of DNA with the same size as a specific gene fragment in a given fraction was obtained from the bulk DNA profile of that fraction. Histograms of the relative amounts of specific fragments and bulk DNA of the same size in each fraction were constructed as shown in Fig. 3, so that the sedimentation of the two could be compared. We have scaled the histograms according to the gradient absorption profile to give the actual sedimentation profiles of the different fragments and obtained results not substantially different from those in Fig. 3 (not shown).

From Fig. 3 it can be seen that ovalbumin gene fragments A-D and collagen gene fragment F sedimented at the same rate, or perhaps slightly faster, than bulk chromatin of equal size, while globin gene fragment E was significantly retarded. A qualitative difference between the sedimentation of globin gene fragment E and collagen gene fragment F can be seen in fig. 1g, in which fragment F is found in a fraction containing bulk

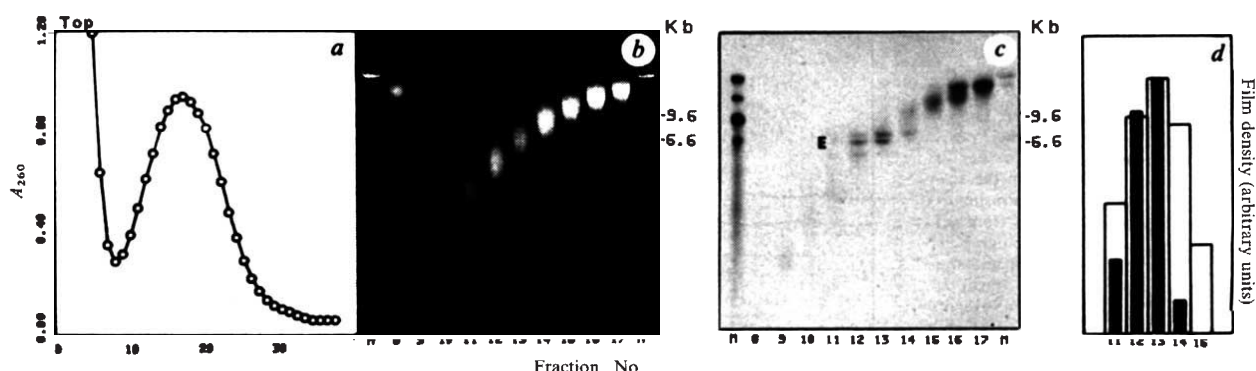


Fig. 4 Globin gene distribution in *EcoRI*-digested spleen chromatin. Globin gene sequences were analysed in *EcoRI* digested spleen chromatin fragments as in Figs 1–3. *a*, Absorbance profile of the sucrose gradient fractions of spleen chromatin fragments. *b*, Fluorescence photograph of the DNA from the indicated sucrose gradient fractions analysed by gel electrophoresis. *c*, Southern blot of the gel in *b* probed for globin sequences and showing the 6.2 kb globin fragment E. *d*, Histogram showing the distribution of the 6.2 kb globin fragment E (solid bars) and bulk DNA of the same size (open bars). Note that, unlike histogram of erythrocyte chromatin shown in Fig. 3(E), the globin sediments with bulk DNA.

Methods: Sodium butyrate was omitted from buffers to ensure a high yield of chromatin; control experiments demonstrated that this omission made no difference to the retardation of globin gene fragments from erythrocyte chromatin (data not shown). A spleen was removed from a White Leghorn chicken, cleaned and washed in PBS with 1 mM EGTA and 0.2 mM PMSF. The spleen tissue was homogenized in buffer A and passed through two layers of surgical gauze. The cells were centrifuged at 2,000g for 5 min and the top half of the pellet (leaving red blood cells below) was collected. Nuclei were prepared as described in Fig. 1. Aliquots of 1 mg ml⁻¹ of nuclei were digested with 20,000 units ml⁻¹ *EcoRI* for 30 min at 37 °C. The nuclei were lysed and the chromatin fractionated and analysed as previously described.

DNA of about the same size (from the repetitive element), whereas fragment E is associated with DNA smaller than itself. Further analysis of the bulk DNA size in each fraction (see Fig. 3 legend) allows us to specify the size of the bulk DNA in chromatin which cosediments with specific fragments. In the peak fractions of both the 9.3 kb ovalbumin fragment D and the 8.3 kb collagen fragment F, the bulk DNA had a number average size of 10.0 kb. By contrast the peak fraction of the 6.2 kb globin fragment E had a bulk DNA size of 4.9 kb. These size comparisons reinforce the qualitative conclusions deduced from Figs 1–3.

Comparable analysis of globin gene fragment E in *EcoRI*-digested chicken spleen chromatin is shown in Fig. 4. Both by direct comparison of the DNA fluorescence pattern with the autoradiograph, and by densitometric analysis, it can be seen that the globin gene fragment E travels with bulk DNA of the same size.

Our previous work⁵ established that chicken erythrocyte chromatin, which in its native state is substantially in the canonical solenoid structure¹⁰, can be reversibly unfolded by the removal and readdition of linker histones and that the sedimentation rate reflects the degree of folding induced by the linker histones. Therefore, the sedimentation rate can be used as a measure of chromatin unfolding and, in the absence of other differences, linker histone depletion. While we cannot categorically exclude that the chromatin in the region of the β -globin gene in chicken erythrocytes differs in some other manner from bulk chromatin, the simplest interpretation of the sedimentation behaviour of the restriction fragments is that the region is selectively depleted of linker histones and, as a result, unfolded. Recent bulk chromatin studies (R. Hannon, unpublished results) indicate that linker histones bind cooperatively to certain regions of chromatin, leaving other regions exposed for polymerase binding and transcription. Nuclease sensitivity studies on isolated chromatin (T.K., unpublished results) also suggest that the globin gene region is selectively depleted of linker histones. The retardation of the globin chromatin fragments from erythrocytes in the sedimentation studies here would

correspond, with reference to our previous calibration⁵, to about a 50% depletion of linker histones relative to bulk chromatin.

Restriction endonuclease digestion has previously been used to advantage for purifying chromatin containing satellite DNA^{11,12} and for visualisation of higher-order structure in nucleolar chromatin along tandem repeats of ribosomal RNA genes¹³. Combined use of chromatin restriction fragments and cloned DNA probes permits extension of physical methods previously applicable only to studies of bulk chromatin⁵ or purified genes^{11–13} to unique genes without prior purification. Determination of sedimentation properties of specific chromatin fragments, exemplified here, is but one of many physical studies made possible by this approach. The structural basis of specific gene expression in eukaryotes should thus be more accessible to study.

We thank Drs G. Felsenfeld and W.I. Wood for λ WesCA8G1, Dr B.W. O'Malley for pOV230 and Drs H. Boedtker and V.E. Tate for p α 2CG71sub68. This work was supported by MRC project grant G979/1094/SA, a Wellcome Trust Senior Fellowship (J.A.) and a King's College Tutorial Studentship (T.K.).

Received 22 April; accepted 19 October 1983.

- Chao, M. V., Mellon, P., Charnay, P., Maniatis, T. & Axel, R. *Cell* **12**, 483–493 (1983).
- Weinbraub, H. & Groudine, M. *Science* **193**, 848–856 (1976).
- McGhee, J., Wood, W. T., Dolan, M., Engel, J. D. & Felsenfeld, G. *Cell* **27**, 45–55 (1981).
- Cartwright, I. L. *et al. CRC Crit. Rev. Biochem.* **13**, 1–86 (1982).
- Allan, J. *et al. J. Cell Biol.* **90**, 279–288 (1981).
- Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).
- Hentschel, C. C. & Tata, J. R. *Dev. Biol.* **65**, 496–507 (1978).
- Gariglio, P., Bellard, M. & Chambon, P. *Nucleic Acids Res.* **9**, 2589–2598 (1981).
- Wood, W. I. & Felsenfeld, G. *J. biol. Chem.* **13**, 7730–7736 (1982).
- Finch, J. T. & Klug, A. *Proc. natn. Acad. Sci. U.S.A.* **73**, 1897–1901 (1976).
- Musich, P. R., Maio, J. & Brown, F. L. *J. molec. Biol.* **117**, 651–677 (1977).
- Fittler, F. & Zachau, H. G. *Nucleic Acids Res.* **7**, 1–17 (1979).
- Pruitt, S. L. & Grainger, R. M. *Cell* **23**, 711–720 (1981).
- Lawson, G. M., Tsai, M.-J. & O'Malley, B. W. *Biochemistry* **19**, 4403–4411 (1980).
- Dolan, M., Sugarman, B. J., Dodgeson, J. B. & Engel, D. *Cell* **24**, 668–677 (1981).
- Wozney, J., Hanahan, D., Tate, V., Boedtker, H. & Doty, P. *Nature* **294**, 129–135 (1981).
- McReynolds, L. A., Catterall, J. & O'Malley, B. W. *Gene* **2**, 217–230 (1978).
- Ginder, G. D., Wood, W. I. & Felsenfeld, G. *J. biol. Chem.* **254**, 8099–8102 (1979).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Rigby, P. W. J., Dickmann, M., Rhodes, C. & Berg, D. *J. molec. Biol.* **113**, 237–251 (1977).
- Denhardt, D. *Biochem. biophys. Res. Commun.* **23**, 641–646 (1966).

Staying neutral on evolution

John Maynard Smith

The Neutral Theory of Molecular Evolution. By Motoo Kimura.
Cambridge University Press: 1983. Pp. 366. £35, \$69.50.

THIS is the book we have been waiting for. Fifteen years ago, the neutral theory was proposed by Kimura, and independently by King and Jukes. In Kimura's formulation, it made two assertions. First, chance events have been more important than natural selection in bringing about evolutionary changes in molecules. Second, existing molecular polymorphisms reflect these chance events, and in most cases are not maintained by balancing selection. How have these assertions stood up to the test of time?

The major implication of the theory is that the rate of evolution of a class of molecules will be high for molecules with few selective constraints, and low for those which are tightly constrained. For strictly neutral changes, the rate of evolution can be shown to be independent of population size, and equal to the "neutral mutation rate": it is therefore proportional to the fraction of all mutations that are selectively neutral. In so far as this fraction is constant for a given class of protein — e.g. haemoglobins — the evolutionary rate for that class will also be constant.

This prediction has been confirmed more fully than even Kimura could have hoped. Constancy of rate is not absolute — it has been shown in some cases to vary more than would be expected by chance — but it is close enough to be widely used as a "molecular clock": in the one case — of the divergence between man and ape — in which there was a disagreement between molecular and palaeontological data, the latter have been "reinterpreted". More important, however, is the negative association between rate and selective constraints. It has been known for some time that protein regions such as fibrinopeptide and insulin C, which are subject to relatively few constraints, evolve much faster than molecules such as cytochrome c, and that surface regions of proteins change faster than internal ones, or than regions close to the active site. More recently, DNA sequencing has revealed that even more rapid evolution occurs at sites where changes do not alter functional proteins. These include synonymous changes in third sites (for which, interestingly, change is as rapid in genes determining proteins such as histone IV, which is virtually unchanging in evolution, as in genes determining rapidly evolving proteins), introns, and "pseudogenes" (i.e. genes derived by duplication from functional genes, but which no longer code for a protein). Even in these regions, not all changes are necessarily neutral. For

example, Kimura discusses the evidence that synonymous changes in third sites are constrained by the relative abundances of the corresponding tRNAs.

My own role in the neutralist debate has been the rather negative one of sitting on the fence, and pointing out from time to time that arguments used on one side or the other are less decisive than their proponents imagine. However, it would be obstinate, in face of the DNA sequence data, to doubt that the most rapid changes occur in regions where they matter least, and occur because they do not matter. I suppose that it is logically possible that changes in third sites and in introns have been neutral, and yet that most amino-acid changes have been selected. But when we bear in mind that it is the least important amino-acid changes that are most frequent, the idea, even if logically possible, seems highly implausible.

There has, I am afraid, been an irrational as well as a rational opposition to Kimura's ideas, particularly in Britain. For example, a recent elementary text about evolution asserted that this theory is wrong because different classes of molecule evolve at different rates. This degree of misunderstanding is hard to excuse. It has, I think, been provoked by the phrase "non-Darwinian evolution". It is therefore worth stressing that Kimura is an enthusiastic Darwinist. He says in his introduction "Classical evolution theory has demonstrated beyond any doubt that the basic mechanism for adaptive evolution is natural selection acting on variations produced by changes in chromosomes and genes". This remark is typical of his outlook.

There are, however, real difficulties with this theory. In particular, the prediction of a constant evolution rate has lost its pristine simplicity and elegance, for two reasons. First, Fitch has shown that the region of a protein that evolves is different in different lineages. Now, so long as we could imagine that in, say, haemoglobin there is a fixed set of residues which are free to change over a fixed set of possible alternatives, it was easy to see why the fraction of all mutations that are neutral remains constant: after Fitch, it is no longer so clear. More serious, it turns out that the rate of evolution is constant relative to absolute time, and not to generation number, although mutation rate is more nearly constant per generation than per year. Kimura recognises that this is a difficulty. Following Ohta, he suggests the following way out. The substitutions that

occur may not be strictly neutral, but may be very slightly deleterious. If so, their chance of establishment is greater in a small than a large population. Animals with long generation times tend to be large, and to have small population sizes. In such animals, a larger proportion of mutations may be "effectively neutral", in the sense that stochastic events are more important than selection in deciding their fate. In consequence, a higher mutation rate to effectively neutral alleles may counter-balance the longer generation time. It will be apparent that, as is only too common in biology, the theory has lost some of its beauty after contact with brute reality — never glad confident morning again! All the same, the theory does explain a lot in a surprisingly simple way.

I have so far said nothing of the second component of the theory, that most molecular polymorphisms are neutral. This aspect has always been harder to test, because the degree of polymorphism to be expected depends critically on population size, in a way that substitution rate does not. Worse still, it is not just present population numbers that matter, but past numbers, which are virtually impossible to estimate (unless, of course, we assume the truth of the theory, and use molecular polymorphism to estimate past numbers). In consequence, the most I can say is that there are features of the data which are consistent with the neutral theory, but nothing that compels one to accept it. Of course, if one accepts that most evolutionary substitutions are neutral, then some existing polymorphisms must be, because every substituted allele was polymorphic once.

So far I have written of Kimura's ideas, rather than of the book itself. The book contains more than an outline of the current state of neutral theory, although of course that is its core. There is a brief history of evolutionary biology, from Lamarck to the present, which contains some fascinating judgements. He praises Fisher's mathematical insight unreservedly, but adds that his conclusion that random drift is unimportant "appears to be almost an anticlimax". He remarks of Ford's famous definition of polymorphism "this was a peculiar definition, for it tacitly presupposes the mechanism involved". He writes of the post-synthesis period "Such terms as integrated coadapted complex . . . were introduced . . . But in my opinion they were more rhetorical than scientific". The book also looks to the future, in an all-too-brief discussion of selfish DNA, unequal crossing over, and "coincidental evolution".

As expected, the book is a work of advocacy. For example, in defending his original argument for the neutral theory (of which an important component was that the genetic load required to produce the observed rate of gene substitution would be excessive) against a criticism of my own, he remarks, perhaps with justification, that

"there are several unrealistic features of Maynard Smith's model". Fair enough. But he does not mention that the figure of 4×10^9 nucleotide pairs in the mammalian genome, which he used to estimate the number of genes and hence of substitutions, even if reasonable at the time, is a substantial over-estimate of the coding DNA, and hence leads to an over-estimate of the load.

There are other occasions on which he overstates his case, but it really does not matter. He is the founder and principal architect of the neutral theory. He is convinced that it explains more about molecular evolution than any alternative theory, and he tells us why. This he does with clarity and vigour. He writes English as if it was his native language, with an elegance that most of us can only envy. He may sometimes reject arguments that I would accept, but he does not ignore facts or arguments which might damage his position. The result is a book which will rank with *The Genetical Theory of Natural Selection* and *The Causes of Evolution* as a milestone in evolutionary biology. It may not be right on every point, but it expresses an original viewpoint which from now on will be part of the way we see the world. □

John Maynard Smith is Professor of Biology at the University of Sussex.

Hazy atmosphere

W.L. Chameides

Atmospheric Chemistry.
Edited by E.D. Goldberg.
Springer-Verlag: 1983. Pp.384.
DM48, \$20.70.

THE realization in the early 1970s that free radicals are a normal component of the Earth's lower atmosphere revolutionized our understanding of atmospheric chemistry and its role in the larger biogeochemical cycling of the elements. However, the advances made in this field over the past decade have yet to be comprehensively covered in a single book. Unfortunately, although informative and enjoyable, *Atmospheric Chemistry* falls short of filling this need.

The book arises from a Dahlem Workshop on the chemistry of the lower atmosphere. Four sub-disciplines were discussed at the meeting: aqueous atmospheric chemistry, historical records in atmospheric composition, biogenic sources of atmospheric constituents and gas-phase tropospheric photochemistry. The book has a section for each of these fields and, in addition, a general introduction by R.M. Garrels. In each section, there are three or four "background" chapters written by individual conference participants, followed by a Group Report summarizing the findings and recommend-

ations of the participants as a whole.

The advantage of such a format is that the diversity of styles and approaches gives the reader a sense of the vitality and excitement of the conference itself. Many of the chapters are excellent. Garrels's description of global chemical cycles as a complex array of interlocking wheels in the introductory chapter is intriguing. The chapter on precipitation-scavenging, liberally laced with poetic quotations and thinly veiled barbs, is characteristic of George Slinn and quite entertaining. In Lovelock's chapter on anoxic environments the reader is treated to some fascinating bits of information on biota; we learn, for instance, why cockroaches secrete octane. In addition there are several very scholarly reviews such as Stuiver's on carbon isotopes and Penkett's on non-methane organics.

However, while the format makes for several very stimulating chapters it prevents the book from succeeding as a whole because it is not consistently organized. For example, in Garrels's introductory chapter it is proposed that atmospheric chemistry is best examined by a consideration of the time constants intrinsic to global chemical cycles. This theme is then dropped in the following chapter and is never brought up again.

Quite often the Group Report chapters would have served better as introductions to the subdisciplines rather than as summaries; for instance in the biospheric sources section two detailed "background" chapters on oxic and anoxic environments are followed by the summary chapter in which the rationale for conceptually dividing the biosphere in this manner is presented. In other cases, the same facts (e.g. the ionic composition of rain) are discussed repeatedly in chapter after chapter; expositions on the "history of atmospheric chemistry" are presented in no less than four chapters. Examples of using specialized terminology without defining the terms or defining them in subsequent chapters can also be cited. Finally there are several rather noticeable holes in the coverage; these include chemical mechanisms in aqueous chemistry and the secular trends in atmospheric trace constituents such as CH_4 and N_2O .

Overall, the book reads as the proceedings of a workshop on atmospheric chemistry rather than a comprehensive text on the subject. The lack of organization between the chapters renders the book of limited value as an introductory treatise for the novice. Furthermore, the rather restricted citation list presented in most chapters prevents its use as a serious reference text. However, the book can be an informative documentary of what must have been a very exciting meeting among some of the more talented scientists now working in this field. □

W.L. Chameides is an Associate Professor of Geophysics at the Georgia Institute of Technology, Atlanta, Georgia.

The good, the bad and the ugly

David L. Hull

**The Eclipse of Darwinism:
Anti-Darwinian Evolution Theories in
the Decades around 1900.**

By Peter J. Bowler.
Johns Hopkins University Press: 1983.
Pp.291. £19.50, \$25.

AS most historians portray it, science is a linear sequence of triumphant discoveries. Occasionally we hear of scientists who were mistaken, just so long as they were mistaken in big ways. But most scientists fail in the most literal sense of this term. They may publish a few papers, but these papers rapidly sink out of sight in the great ocean of scientific publications. In most cases, publishing a paper is equivalent to throwing it away. Historians of science complain that chronicling only highly visible scientists at the expense of their more obscure fellow labourers is as misleading as telling the story of biological evolution without mentioning the vast majority of species that become extinct. The trouble is that few people want to read about experiments that failed, research that led nowhere, and ideas that were so inconsequential that no one was interested. Historians have written occasional histories of nobodies, but no one seems willing to read them including the nobodies.

Histories of Darwinism inevitably begin with Darwin's roots in uniformitarian geology and natural history, through Darwin and his fellow Darwinians, the advances of August Weismann and the rise of Mendelian genetics, to the New Synthesis. In *The Eclipse of Darwinism*, Peter Bowler takes just the opposite tack. He traces the anti-Darwinians from the earliest theistic teleologists, through the neo-Lamarckians and orthogenecists to Hugo de Vries and his mutation theory. Bowler treats his readers to a fascinating panoply of largely unfamiliar entities and processes — entities such as archetypes, plastidules, the idioplasm, mnemes and hopeful monsters, and processes such as intracellular pangenesis, kinetogenesis, orthogenesis, aristogenesis and racial senility. The names of the scientists whom Bowler discusses are somewhat more familiar — Louis Agassiz, St George Jackson Mivart, Henri Bergson, Richard Semon, Paul Kammerer, Theodor Eimer, Henry Fairfield Osborn, Alpheus Hyatt, Edwin Drinker Cope, Hugo de Vries, William Bateson, Richard Goldschmidt and D'Arcy Thompson. Terming all these scientists "losers" in the same sense is hardly accurate. Many of them made important contributions to science, while others at least had their day in the sun of scientific acceptance. They do all have one

thing in common however — on those issues where they came into conflict with the Darwinians, they eventually lost out. They may have been losers, but at least some of them were big losers.

As Bowler reminds us, Darwinism originally included Lamarckian inheritance as a subsidiary mechanism. Both Herbert Spencer and Ernst Haeckel were confirmed Lamarckians. Not until August Weismann purged Darwinism of the inheritance of acquired characters did Darwinians and Lamarckians become opponents, much to the dismay of such pluralist Darwinians as George John Romanes. As Bowler also notes, we tend to forget that early Mendelians were just as opposed to Lamarckian and Darwinian theories of evolution. During the first decade of the century, the Mendelians became the gatekeepers of biology. Any theory that could not come up to their standards was put in jeopardy. Lamarckism was defeated not by the Darwinians but by the Mendelians. The Lamarckians lost because they refused to divorce their theory of evolution from embryology. The Darwinians survived because they were willing, at least for the moment, to ignore embryology and ally themselves with the overly simple views and extremely successful methods of Mendelian genetics. Later they would put organisms back into evolution.

One factor that Bowler suggests to explain the eclipse of Darwinism is the polarization produced by Weismann's "dogmatic" emphasis on natural selection to the exclusion of all other evolutionary mechanisms. The inflexibility of the resulting neo-Darwinians gave birth to the neo-Lamarckians and guaranteed that they would be enemies. Bowler tells a parallel story for dogmatism about gradual (continuous) versus saltative (discontinuous) evolution. Bowler's choice of words implies that flexibility is inherently superior to dogmatism, but both are strategies and as such are sometimes successful, sometimes not. In science as in biological evolution, rigidity is not always a vice and flexibility not always a virtue.

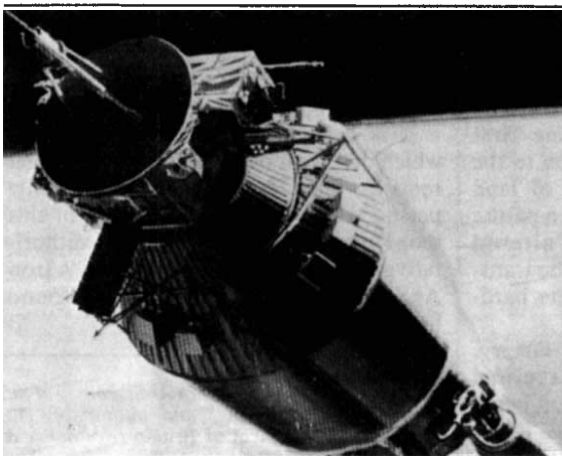
Bowler's carefully documented book gives precious little support to those who

think that great scientific theories spring full blown from the brains of great scientists and remain essentially unchanged thereafter. Although Darwinism underwent periods of retrenchment when certain Darwinians pruned the luxuriant pluralism of their theory, just about every imaginable view has been part of Darwinism at one time or another. Although Goldschmidt's "hopeful monsters" were ridiculed by such founders of the Synthetic Theory as G.G. Simpson and Ernst Mayr, the "hopeful populations" of Mayr, Niles Eldredge and Stephen Jay Gould are rapidly becoming part and parcel of the *new* New Synthesis.

Bowler also offers scant solace to the more extreme social determinists. Any scientific view seems capable of justifying any social policy. For example, Spencer was both a Lamarckian and a strong advocate of *laissez-faire* individualism. Bowler concludes that those opponents of modern selection theory who "dismiss it as an expression of blind materialism or capitalist ideology should pause for a moment to think that if it were not for the triumph of the modern synthesis, their own alternative might still be burdened with equally distasteful implications" (p.221).

Construing science as nothing but a series of success stories is surely misleading, but Bowler, in attempting to compensate, also tends to bias the picture. Scientists are frequently quite stubborn. If they were not, they could not begin to overcome the tremendous inertia which they frequently confront. When they turn out to be right, we admire their stubbornness. Bowler's litany of stubborn adherence to views that we now take to be mistaken tends to get a bit depressing. Science moves so fast that most of the scientists whom Bowler discusses lived long enough to see the scientific community move out from under them. The resulting bitterness and petulance is also none too attractive, but Peter Bowler's antidote to the relentless exaltation of traditional histories of Darwinism is as necessary as it is sometimes distasteful. □

David L. Hull is Distinguished Professor of Philosophy at the University of Wisconsin, Milwaukee.



GLIMPSE of 1986 — an artist's impression of the International Solar Polar Mission spacecraft (which is now completed) atop a Centaur upper stage after its deployment from the Space Shuttle. The illustration is taken from a new publication of the European Space Agency, *The International Solar Polar Mission — Its Scientific Investigations*, which previews the work to be carried out by the ISPM. The publication (ESA SP-1050) is available from Scientific and Technical Publications Branch, ESTEC, Noordwijk, The Netherlands. Price is 175FF.

Full of Fourier

H. Lipson

Fourier Optics: An Introduction.

By E. G. Steward.

Ellis Horwood/Wiley: 1983. Pp.184.

Hbk £15, \$25.95; pbk £7.95, \$13.75.

OVER the past three decades there has been a revolution in physical optics. In the 1950s it was generally held that it was a finished subject: we knew all that we needed to know and while optics was still a necessary part of a physics curriculum, no creative ideas could be expected to arise in it.

However, developments were taking place. The Dutch physicist, Fritz Zernike, deliberately chose to devote himself to optics rather than to the more fashionable topics; he generalized the concept of coherence and, in 1935, invented the phase-contrast microscope. In 1960 Maiman made the first laser, which astounded the scientific world, giving a beam with practically infinite coherence. This finding was seized upon by Leith and Upatnieks in 1962 to make workable the hologram idea introduced by Gabor in 1942. Optics had been rejuvenated and now anyone who wishes to consider himself a physicist must know something of these new ideas.

This book will be extremely helpful in introducing these approaches. The name Fourier dominates the contents, illustrating in addition how the different branches of physics are interdependent; Fourier, of course, was concerned with heat transmission, not optics, but now his mathematical idea is used in almost every branch of physics.

I think that the book succeeds very well. The style is pleasantly informal but nonetheless quite explicit. The author has thought deeply about the presentation of the subject. Fourier transformation and convolution are clearly described and are applied to explanations of optical imaging and processing, subjects that have made great strides in recent years. In the final chapter, which is concerned with instrumentation, the name of Michelson looms large; the use of Fourier methods in the interpretation of the patterns from the spectral interferometer is perhaps one of the most surprising outcomes of Fourier's ideas.

The only criticism I have of this section is that there is no clear explanation of the way in which fringes are formed in the interferometer, and the problem of localization of fringes — always difficult to explain to students — is not dealt with at all.

In all, however, the book is very readable and will be most useful to anyone who wishes to keep up to date with modern optical ideas. □

H. Lipson is Emeritus Professor of Physics at the University of Manchester Institute of Science and Technology.

Behavioural ecology and the female

Tim Clutton-Brock

Social Behavior of Female Vertebrates.
Edited by Samuel Wasser.
Academic: 1983. Pp.395. \$35, £23.20.

MORE than most other areas of behavioural ecology, studies of social behaviour have suffered from the "Could Be" school of adaptationist interpretation. One of the principal accomplishments of this approach is the ability to transform introductory statements — "the pink coloration typical of flamingoes could be a form of camouflage which allows them to merge into tropical sunsets thus reducing the probability of twilight predation" — into firmer assertions without recourse to direct evidence by the time the Discussion section has been reached. ("By merging into the glaring highlights of tropical sunsets and thus reducing nocturnal mortality, flamingoes can afford to defer reproduction until the onset of senescence and thereby. . .").

The "Could Be" approach is used to float novel hypotheses, buoyed up by heavily selected references. Most of us have contributed to it at some stage, but the standards of evidence in behavioural ecology are rising fast and the general trend is towards the critical evaluation of existing hypotheses and away from the production of new theory. With one or two exceptions, the thirteen chapters of *Social Behavior of Female Vertebrates* reflect this development. The book includes thoughtful reviews of the evolution of polyandry (Erckmann) and the role of bird song in

mate choice (Payne), detailed investigations of mate choice in sculpins (Brown and Downhower), altruism in coatis and elephants (Russell and Dublin) and an outstanding analysis of cooperation among Acorn woodpeckers (Koenig, Mumme and Pitelka). Two studies of female strategies in humans suffer from their proximity to the animal papers for this emphasizes the contrasting quality of data available to zoologists and anthropologists.

But why female vertebrates? In the two opening chapters, Hrdy, Williams, Wasser and Waterhouse point to a male bias both in the orientation of early behavioural research and in the gender of the researchers. More recently, evidence of the extent to which breeding success varies among females has increased and the proportion of female researchers has grown. This, they argue, has been responsible for increasing interest in female reproductive strategies which has quickly shown how inaccurate was the early view of females as the passive subjects of male manipulation.

In places, their history of the sex bias is exaggerated and its attribution to masculine bias a shade too glib. While male bias has certainly been involved, there are reasons why it can be harder to collect detailed data on female breeding strategies than on those of males — female mammals are often less visible than males and their longer breeding lifespans and reduced variance in seasonal success complicate the study of breeding strategies.

However, it is not the past bias against studying females which justifies the focus of the book, but the fact that major insights can be achieved by considering the behaviour of the two sexes separately. The book makes an important contribution in synthesizing recent developments in the study of female behaviour and the next step, as Wasser points out, is to investigate how male and female strategies interact. □

Tim Clutton-Brock is in the Large Animal Research Group of the Department of Zoology, University of Cambridge.

Manna to manure

Richard Bradley

The Farming of Prehistoric Britain.
By P.J. Fowler.
Cambridge University Press: 1983.
Pp.246. £7.50, \$14.95.

SEVERAL years ago the British Broadcasting Corporation persuaded a group of volunteers to build and operate an Iron Age village. The experiment was not a success: some of their animals escaped, crops withered in the fields and replicas of prehistoric implements were difficult to use. To some extent these problems arose because archaeologists themselves were so uncertain how prehistoric Britain was farmed, but the greatest obstacle of all was the lack of a basic textbook setting out what was known and what seemed likely to have happened. Fowler's new book fills that need.

Its subject matter is vast. As late as 4,000 BC, when farming had developed in parts of Europe, the British economy was based on hunting, gathering and fishing. By the Roman Conquest the native population was exporting grain to the Continent. Fowler's book traces the development of prehistoric agriculture from the first clearance of the natural vegetation to the growth of complicated systems of land management. The balance between nature and the human population altered drastically during this time with the transformation of hunter-gatherers into hard-pressed farmers.

Fowler's book has an unusual history, although it hardly shows it. It is a revised and expanded version of his section in the *Agrarian History of England and Wales* (Cambridge University Press, 1981), an

uneven and expensive review of prehistoric farming. The revised text swallows up the fields discussed by other contributors to that volume and covers a longer period of time. The new book — and it is a new book — consists of the original text, a substantial series of additions, amendments and a number of still more recent footnotes. Despite these piecemeal changes, the end result is lucid, lively and convincing. It is less surprising that it is so up to date. Unfortunately, the figures are taken from the earlier volume and in some cases could have been replaced by more recent or more apposite examples.

The book does have certain weaknesses. The major emphasis is upon ecological factors; developments in prehistoric agriculture are caused mainly by population pressure and climatic change. A subsidiary influence may be immigration from abroad. The case is not entirely clear-cut and still more could have been made of internal social factors. For example, it seems likely that by the Iron Age grain supplies were being mobilized as tribute, and this must have influenced farming practice. Again, the text deals mainly with the Bronze and Iron Ages and too little space is devoted to the earlier development of agriculture. Perhaps there is room for a longer second edition.

It is easy to suggest other questions which deserve more attention, but the fact remains that Peter Fowler is the only person who has produced a book of this scope, and he has done it well. It is authoritative but it is never dull. The BBC's Iron Age Family would certainly have found it useful. □

Richard Bradley is Reader in Archaeology at the University of Reading, and author of The Prehistoric Settlement of Britain (Routledge & Kegan Paul, 1978).

New in paperback

NOTABLE among recent paperback editions of previously published books is a reprint of Steven Weinberg's classic, *The First Three Minutes*.

The book was first published in Britain in 1977, and was reviewed by John Barrow in *Nature* 267, 291; 1977. The reprint includes a brief Afterword by Weinberg, which brings the story up to date, and is published by Flamingo (an imprint of Fontana Books). Price is £2.50.

William Broad and Nicholas Wade's *Betrayers of the Truth: Fraud and Deceit in the Halls of Science*, which appeared earlier this year, is a more contentious book and had a mixed reception from reviewers. In *Nature*, however, Walter Gratzer called it "most captivating" (302, 774; 1983). A paperback edition has been issued by Touchstone Books, an imprint of Simon & Schuster, at \$6.95.

Among the best of the many books thrown up by the Creationism controversy was Philip Kitcher's *Abusing Science: The Case Against Creationism* (MIT Press, 1982). The Open University Press (12 Coffridge Close, Stony Stratford, Milton Keynes MK11 1BY, England) has recently published the book in paperback, price £6.95. John Habgood, now Archbishop of York, reviewed *Abusing Science* in *Nature* 300, 118; 1982.

Awards

The Dr Tomalla foundation in Vaduz, Liechtenstein, has awarded its first prize for outstanding contributions to gravitation and cosmology. The SF 50,000 prizes have been awarded to **Dr Subrahmanyan Chandrasekhar**, professor of astrophysics at the University of Chicago and **Dr Andrej Dmitrievich Sakharov**, member of the Soviet Academy of Sciences. Dr Sakharov is currently living in Gorki, USSR.

Professor A.C. Wiin-Nielsen, secretary general of the World Meteorological Organization, has been awarded the Wihuri International prize. The prize is awarded for contributions to cultural and economic development.

The British Society for Drug Research has given its award for drug discovery to **F. Peter Doyle**. Mr Doyle will give a lecture on 'Science and the Pharmaceutical Industry: a Pot-Pourri of Facts, Thoughts and Prejudices' at the presentation ceremony which will take place on 16 December 1983 at the School of Pharmacy, Brunswick Square, London WC1, UK.

The US National Academy of Engineering has recently made two awards. The 1983 Founders award has been presented to **Dr Harold Edgerton**, institute professor emeritus of the Massachusetts Institute of Technology, for his work on ultra-high speed photography. The Arthur M. Bueche award for statesmanship in science and technology has been given to **Dr Simon Ramo** for his work on science policy.

Nominations are invited for the Marconi international fellowship award in communication science and technology and their involvement in human welfare. The award, which consists of \$35,000, will be presented in 1985 and nominations must be received by 30 January 1984. Further information is available from the Marconi International Fellowship, Polytechnic Institute of New York, 333 Jay Street, Brooklyn, New York 11201, USA.

The British Institution of Mining and Metallurgy invites applications for the following grants and fellowships. The Bosworth Smith trust fund gives £2,000 for the assistance of postgraduate research in metal mining, non-ferrous extraction metallurgy or mineral dressing. Grants may be given towards working expenses, the cost of visits to mines and plants and the purchase of equipment. Stanley Elmore fellowships (£1,500-£6,000) are awarded for research into all branches of extractive metallurgy and mineral processing. The fellowships are tenable at universities in the

UK. In addition two Atlas Copco travel bursaries are being offered to younger mining graduates for study tours of Swedish mines; one bursary is open to engineers in any country who have at least three years' practical mining experience, and the other is open to an engineer who is studying at a British university and who has a minimum of one year's practical experience. Applications are also invited for awards from the income of the G. Vernon Hobson bequest (£1,000). These may be made for travel, research or other approved projects. Finally the Edgar Pam fellowship (£1,300) will be awarded in October 1984 for postgraduate study in subjects within the Institution's fields of interest. The fellowship is open to residents of Australia, Canada, New Zealand, South Africa and the UK. Application forms and further information may be obtained from The Secretary, The Institution of Mining and Metallurgy, 44 Portland Place, London W1N 4BR, UK.

Applications are invited for the Curt P. Richter prize in psychoneuroendocrinology which has been established by the Institute of Clinical Pharmacology of Dublin, Ireland for the International Society of Psychoneuroendocrinology. The prize, which is to be awarded annually, consists of \$1,000. It will be given for the best essay or manuscript describing original research or for a review including original research. The piece must be submitted by a scientist or physician who is under 40 on 1 January 1984. The paper should be compatible with publication in *Psychoneuroendocrinology*. Manuscripts should be submitted to Dr Gerhard Langer, Associate Professor of Psychiatry, Department of Psychiatry, University of Vienna, Lazarettgasse 14, A-1097 Vienna, Austria.

Applications are invited for Lister Institute research fellowships, of which up to five will be available from October 1984. Each fellowship will be tenable for five years. Candidates must be postdoctoral researchers in the biomedical field and aged under 34 years. They must show evidence of independent research, including at least two years' postdoctoral research, and they must have an interest in a scientific aspect of preventive medicine. Special arrangements will be considered for those clinicians who are required to undertake further training for accreditation in their speciality. If the fellow has temporarily vacated a permanent post, it is expected that the employing institution will appoint a temporary replacement during the period of the fellowship. The salary, payable by the employing institution and reimbursed by the Institute, will normally be on university clinical or non-clinical scales. An

allowance for consumables and minor equipment is also payable. Further details are available from The Secretary, Lister Institute of Preventive Medicine, Elstree, Hertfordshire WD6 3AT, UK.

Appointments

Professor John Kingman, chairman of the Science and Engineering Research Council, has been elected vice-chancellor of the University of Bristol, UK. He is to succeed Sir Alec Merrison. Professor Kingman will take up his appointment on 1 September 1985. In addition, **Mr D.T. Ulph**, currently of University College, London, has been appointed professor of economics and **Mr A.D. Cheshier**, currently of the University of Birmingham, has been appointed professor of econometrics at Bristol.

Dr George Hill, formerly an associate professor in the department of pathology, Colorado State University, has been appointed professor and director of the division of biomedical sciences, Meharry Medical College, Nashville, Tennessee, USA.

Professor John Newsom-Davis and **Professor Peter Morris** have been appointed to the Medical Research Council, UK.

Dr Brian Mahy has been appointed director of the British Agricultural Research Council's Animal Virus Research Institute. He will succeed Dr R.F. Sellers who is to retire at the end of August 1984.

Professor E. Christopher Zeeman has been appointed Professor of Mathematics in the Royal Institution.

The London School of Hygiene and Tropical Medicine has announced that **Dr G.A.T. Targett** has been made professor of immunology of protozoal diseases by the University of London.

Professor Peter Frank Stott, currently director general of the water council, has been appointed to the Kevin Nash chair of civil engineering at King's College, London, UK.

Dr Charles Michel has been appointed to the chair of physiology tenable at St Mary's Hospital Medical School, London, UK.

Miscellaneous

The British Agricultural Research Council (ARC) has formally adopted the title Agricultural and Food Research Council (AFRC).

The Royal Society of Chemistry has launched a new monthly brochure, *Chemical Hazards in Industry*. The brochure is intended to give information on all aspects of health and safety for those working in chemical plants and similar facilities. It will assimilate all recently-published relevant information and will give reference lists for further investigation.

The Medical Advisory Committee, set up to advise the Health and Safety Commission on aspects of occupational health, has been given a further three-year term of office. The committee's terms of reference are still to consider and advise the commission on biomedical aspects of occupational health, including the identification of health hazards, biological monitoring, epidemiological studies, mental health and rehabilitation. It will also cover other associated matters referred to it by the commission or the Health and Safety Executive. **Dr Tim Carter**, the director of medical services of the Health and Safety Executive, is the new chairman of the committee.

Fifty visiting fellows will be elected to take part in an inter-disciplinary project devoted to all aspects of the intellectual and artistic life in Scotland during the period of roughly 1750–1800. The project is being mounted by the Institute for Advanced Studies in Humanities of the University of Edinburgh, UK. The fellowships will be tenable for periods of one to six months. Learned societies will hold their meetings in Edinburgh during the project and a number of professional and civic bodies will organize lectures and exhibitions associated with the project. Further information is available from The Directors of IPSE, 17 Buccleuch Place, Edinburgh EH8 9LN, UK.

Meetings

23–25 January, **Conference on the Biochemistry of the Reproductive Years**, Tucson, Arizona (Ms Marjorie Hack, 355 St Paul's Avenue, Staten Island, New York 10304, USA).

14 February, **Symposium on Cognitive Science**, Tel Aviv (Ms Edna Margalit, The Colloquium Coordinator, PO Box 4070, Jerusalem, Israel).

19–22 February, **Congress for Recombinant DNA Research**, San Diego, California (Dr Steven Nordeen, Department of Microbiology and Immunology, 804 FLOB, University of North Carolina, Chapel Hill, North Carolina 27514, USA).

19–23 February, **NATO Advanced Research Workshop on Mechanisms of Secondary Brain Damage**, Vipiteno, Italy (Professor A. Baethmann, Institut für Chirurgische Forschung, Ludwig-Maximilians-Universität München, Klinikum Grosshadern, Marchioninistrasse 15, PO Box 70 12 60, D-8000 München 70, FRG).

20–24 February, **Workshop on Lasers in Medicine**, University of Essex, UK (The Liaison Officer, Laser Workshop Courses, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK).

22 February, **Congress for the Automation Scale-Up and the Economics of Biological Process Engineering**, San Diego, California (Automation, Scale-Up, c/o Scherago Associates, Inc., 1515 Broadway, New York, New York 10036, USA).

6–9 March, **Oceanology International**, Brighton, UK (Spearhead Exhibitions Ltd, Rowe House, 55–59 Fife Road, Kingston-upon-Thames, Surrey KT1 1TA, UK).

19 March–6 April, **International Training Course on Immunogenetics**, Bombay, India (For Indian applicants; Dr M.G. Deo, Research Director, Cancer Research Institute, Bombay 400012, India; for non-Indian applicants: Dr M. Weigert, Institute for Cancer Research, 7701 Burholme Avenue, Philadelphia, Philadelphia 19111, USA).

22 March, **Debye Symposium**, Cornell University, USA (Earl Peters, Executive Director, Department of Chemistry, Cornell University, Baker Laboratory, Ithaca, New York 14853, USA).

26–28 March, **European Congress on Fluid Machinery for the Oil, Petrochemical and Related Industries**, The Hague, The Netherlands (Press and Information Officer, The Institution of Mechanical Engineers, 1 Birdcage Walk, Westminster, London SW1H 9JJ, UK).

27–28 March, **Principles of Modern Reversed Phase Chromatography**, West Berlin (Dr I. Molnar, Institut für angewandte Chromatographie, Blücherstrasse 22, D-100 Berlin 61, FRG).

4–5 April, **Annual Meeting of the National Council on Radiation Protection and Measurements**, Washington DC, USA (Mr Roger Ney, NCRP, 7910 Woodmont Avenue, Suite 1016, Bethesda, Maryland 20814, USA).

9–11 April, **Sea Grant Lecture and Seminar on the Biotechnology of Marine Polysaccharides**, Massachusetts Institute of Technology (Ms Therese Henderson, MIT Sea Grant Information Centre, 77 Massachusetts Avenue, Building E38-302, Cambridge, Massachusetts, USA).

9–12 April, **Colloquium of the International Astronomical Union on Very Large Telescopes, their Instrumentation and Programmes**, Garching, FRG (IAU Colloquium No. 79, Ms Christina Stoffer, European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-8046 Garching bei München, FRG).

9–13 April, **International Symposium on Risks and Benefits of Energy Systems**, Jülich, FRG (Mr Hans-Friedrich Meyer, Division of Public Information, International Atomic Agency, Wagramerstrasse 5, PO Box 100, A-1400 Vienna, Austria).

10–13 April, **Meeting of the British Mycological Society**, Manchester, UK (Dr David

Moore, Department of Botany, The University, Manchester M13 9PL, UK).

16–19 April, **International Symposium on Remote Sensing of the Environment — Remote Sensing for Exploration Geology**, Colorado Springs, USA (Dr Robert Rogers, Environmental Research Institute of Michigan, PO Box 8618, Ann Arbor, Michigan 48107, USA).

30 April–2 May, **Colloquium on Protides of the Biological Fluids**, Brussels, Belgium (Colloquium 'Protides of the Biological Fluids' Secretariate, Institute for Medical Biology, Alsebergsesteenweg 196 Chaussée d'Alseberg, B-1180 Brussels, Belgium).

2–5 May, **Symposium on Scientific Methodologies Applied to Works of Art**, Florence, Italy (Istituto Guido Donegani SpA, Segreteria Tecnica, Via G. Fauser 4, 28100 Novaro NO, Italy).

2–4 May, **Symposium on Industrial Activity in Space**, Stresa, Italy (Eurospace, Mrs Rosy Plet, Public Relations, 16 bis, Avenue Bosquet, 75007 Paris, France).

7–9 May, **International Conference on Occupational Ergonomics**, Toronto, Canada (The Conference Office, Human Factors Association of Canada, PO Box 1085, Station B, Rexdale, Ontario, Canada M9V 2B3).

8 May, **International Symposium on Crop Protection**, State University, Gent, Belgium (The Secretary, International Symposium on Crop Protection, Faculteit van de Landbouwwetenschappen, Coupure links 533, B-9000, Gent, Belgium).

8–11 May, **International Seminar on Anaerobic Digestion and Carbohydrate Hydrolysis of Waste**, Luxembourg (Mr P.P. Rotondo, Commission of the European Communities, DG XIII-A2, PO Box 1907, L-2920 Luxembourg).

14–18 May, **Annual Symposium on Safeguards and Nuclear Material Management**, Venice, Italy (L. Stanchi, CEC-JRC, 1-21020 Ispra (Varese), Italy).

16–24 May, **International Symposium on the Chemistry of the Mediterranean**, Primosten, Yugoslavia (Centre for Marine Research-Zagreb, The Rudjer Bošković Institute, PO Box 1016, 410001 Zagreb, Yugoslavia).

21–25 May, **International Conference on Reliability and Maintainability**, Perros-Guirec, France (Mr Roland Goarin, President du comité de programme, 4ème colloque international sur la fiabilité et la maintenabilité, Centre de Fiabilité, CNET-LAB/ICM, PO Box 40, 22301 Lannion, France).

23–26 May, **International Symposium on Endocoids**, Fort Worth, Texas (Dr Frank LaBella, Department of Pharmacology, University of Manitoba, College of Medicine, Winnipeg, Manitoba, Canada R3Z 0W3).

23–25 May, **Conference on Opioid Peptides in Periphery**, Rome, Italy (Organizing Secretariat, APE, Via Giorgio Vasari, 4-00196 Rome, Italy).